

Complexity

# Complexity Problems Handled by Advanced Computer Simulation Technology in Smart Cities 2020

Lead Guest Editor: Zhihan Lv

Guest Editors: Kaoru Ota, Jaime Lloret, Wei Xiang, and Paolo Bellavista





---

**Complexity Problems Handled by Advanced  
Computer Simulation Technology in Smart  
Cities 2020**

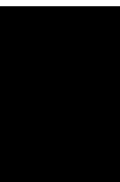
Complexity

---

**Complexity Problems Handled by  
Advanced Computer Simulation  
Technology in Smart Cities 2020**

Lead Guest Editor: Zhihan Lv

Guest Editors: Kaoru Ota, Jaime Lloret, Wei Xiang,  
and Paolo Bellavista



---

Copyright © 2022 Hindawi Limited. All rights reserved.

This is a special issue published in "Complexity." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

# Chief Editor

Hiroki Sayama, USA

## Editorial Board

Oveis Abedinia, Kazakhstan  
José Ángel Acosta, Spain  
Andrew Adamatzky, United Kingdom  
Marcus Aguiar, Brazil  
Carlos Aguilar-Ibanez, Mexico  
Mojtaba Ahmadi Khanesar, United Kingdom  
Tarek Ahmed-Ali, France  
Alex Alexandridis, Greece  
Basil M. Al-Hadithi, Spain  
Diego R. Amancio, Brazil  
Maia Angelova, Australia  
David Arroyo, Spain  
Tomaso Aste, United Kingdom  
George Bassel, United Kingdom  
Abdellatif Ben Makhlouf, Saudi Arabia  
Rosa M. Benito, Spain  
Johan Bollen, USA  
Erik M. Boltt, USA  
Mohamed Boutayeb, France  
Dirk Brockmann, Germany  
Átila Bueno, Brazil  
Seth Bullock, United Kingdom  
Ning Cai, China  
Eric Campos, Mexico  
Anirban Chakraborti, India  
Émile J. L. Chappin, The Netherlands  
Chih-Chiang Chen, Taiwan  
Yu-Wang Chen, United Kingdom  
Diyi Chen, China  
Siew Ann Cheong, Singapore  
Hocine Cherifi, France  
Matteo Chinazzi, USA  
Giulio Cimini, Italy  
Danilo Comminiello, Italy  
Roger Cremades, The Netherlands  
Salvatore Cuomo, Italy  
Sergey Dashkovskiy, Germany  
Manlio De Domenico, Italy  
Pietro De Lellis, Italy  
Albert Diaz-Guilera, Spain  
Thach Ngoc Dinh, France  
James A. Dixon, USA  
Alan Dorin, Australia

Sheng Du, China  
Jordi Duch, Spain  
Marcio Eisencraft, Brazil  
Mondher Farza, France  
Guilherme Ferraz de Arruda, Italy  
Giacomo Fiumara, Italy  
Thierry Floquet, France  
José Manuel Galán, Spain  
Lucia Valentina Gambuzza, Italy  
Harish Garg, India  
Georgi Yordanov Georgiev, USA  
Carlos Gershenson, Mexico  
Peter Giesl, United Kingdom  
Sergio Gómez, Spain  
Lingzhong Guo, United Kingdom  
Xiangui Guo, China  
Sigurdur F. Hafstein, Iceland  
Zakia Hammouch, Morocco  
Chittaranjan Hens, India  
Giacomo Innocenti, Italy  
Sarangapani Jagannathan, USA  
Mahdi Jalili, Australia  
Marco Javarone, United Kingdom  
Peng Ji, China  
Hang-Hyun Jo, Republic of Korea  
Jeffrey H. Johnson, United Kingdom  
Fariba Karimi, Austria  
Mohammad Hassan Khooban, Denmark  
Toshikazu Kuniya, Japan  
Jurgen Kurths, Germany  
C. H. Lai, Singapore  
Guang Li, United Kingdom  
Qingdu Li, China  
Fredrik Liljeros, Sweden  
May T. Lim, Philippines  
Xinzhi Liu, Canada  
Chongyang Liu, China  
Xiaoping Liu, Canada  
Joseph T. Lizier, Australia  
Francesco Lo Iudice, Italy  
Rosa M. Lopez Gutierrez, Mexico  
Vittorio Loreto, Italy  
Nishant Malik, USA  
Rosario Nunzio Mantegna, Italy

Inés P. Mariño, Spain  
Eulalia Martínez, Spain  
André C. R. Martins, Brazil  
Rossana Mastrandrea, Italy  
Naoki Masuda, USA  
Jose F. Mendes, Portugal  
Ronaldo Parente De Menezes, United Kingdom  
Fanlin Meng, United Kingdom  
Marcelo Messias, Brazil  
Ana Meštrović, Croatia  
Ali Minai, USA  
saleh mobayen, Taiwan, R.O.C., Iran  
Osnat (Ossi) Mokryn, Israel  
Christopher P. Monterola, Philippines  
Marcin Mrugalski, Poland  
Jesus Manuel Muñoz-Pacheco, Mexico  
Roberto Natella, Italy  
Chrystopher L. Nehaniv, Canada  
Sing Kiong Nguang, New Zealand  
Vincenzo Nicosia, United Kingdom  
Irene Otero-Muras, Spain  
Yongping Pan, Singapore  
Andreas Pape, USA  
María Pereda, Spain  
Nicola Perra, United Kingdom  
Giovanni Petri, Italy  
Cornelio Posadas-Castillo, Mexico  
Mahardhika Pratama, Singapore  
Karthikeyan Rajagopal, India  
Gabriel Ramos-Fernández, Mexico  
Andrea Rapisarda, Italy  
Andrea Roli, Italy  
Céline Rozenblat, Switzerland  
Daniele Salvati, Italy  
M. San Miguel, Spain  
Matilde Santos, Spain  
Fabio Saracco, Italy  
Antonio Scala, Italy  
Samuel V. Scarpino, USA  
Enzo Pasquale Scilingo, Italy  
Dan Selișteanu, Romania  
Saray Shai, USA  
Wen-Long Shang, China  
Roberta Sinatra, Italy  
Dimitrios Stamovlasis, Greece  
Samir Suweis, Italy

Misako Takayasu, Japan  
Ana Teixeira de Melo, Portugal  
Vito Trianni, Italy  
Shahadat Uddin, Australia  
Gaetano Valenza, Italy  
Jose C. Valverde, Spain  
Sander E. Van Der Leeuw, USA  
Alejandro F. Villaverde, Spain  
Dimitri Volchenkov, USA  
Christos Volos, Greece  
Qingling Wang, China  
Wenqin Wang, China  
Zidong Wang, United Kingdom  
Yong Xu, China  
Honglei Xu, Australia  
Xiao-An Yan, China  
Xingang Yan, United Kingdom  
Zhile Yang, China  
Baris Yuce, United Kingdom  
Massimiliano Zanin, Spain  
Hassan Zargarzadeh, USA  
Hector Zenil, United Kingdom  
Xianming Zhang, Australia  
Zhen Zhang, China  
Fengtai Zhang, China  
Rongqing Zhang, China  
Xiaopeng Zhao, USA  
Tao Zhou, China  
Asim Zia, USA

# Contents

## **Complexity Problems Handled by Advanced Computer Simulation Technology in Smart Cities 2021**

Zhihan Lv , Kaoru Ota, Jaime Lloret , Wei Xiang, and Paolo Bellavista 

Editorial (3 pages), Article ID 9847249, Volume 2022 (2022)


## **Analysis of Feature Extraction and Anti-Interference of Face Image under Deep Reconstruction**

### **Network Algorithm**

Jin Yang , Yuxuan Zhao , Shihao Yang , Xinxin Kang , Xinyan Cao, and Xixin Cao



Research Article (15 pages), Article ID 8391973, Volume 2021 (2021)

## **E-Commerce Enterprise Supply Chain Cost Control under the Background of Big Data**

Haijun Mao and Long Chen 


Research Article (11 pages), Article ID 6653213, Volume 2021 (2021)

## **Complex System of Vertical Baduanjin Lifting Motion Sensing Recognition under the Background of Big Data**

Yan Zhang, M. M. Kamruzzaman , and Lu Feng 

Research Article (10 pages), Article ID 6690606, Volume 2021 (2021)

## **Earthquake Disaster Rescue Model Based on Complex Adaptive System Theory**

Fujiang Chen , Jingang Liu, and Junying Chen

Research Article (12 pages), Article ID 6655574, Volume 2021 (2021)

## **Construction of a Complex System Based on Big Data for the Intelligent Service System of Youth Physical Health**

Xinwen Li, Chao Song, Christine A. Rochester, and Chaobing Yan 

Research Article (12 pages), Article ID 6635346, Volume 2021 (2021)

## **Flood Simulation Analysis of the Biliu River Basin Based on the MIKE Model**

Qi Liu , Dianwu Wang , Yulong Zhang, and Li Wang


Research Article (10 pages), Article ID 8827046, Volume 2021 (2021)

## **Cluster Coordination between High-speed Rail Transportation Hub Construction and Regional Economy Based on Big Data**

Liang Zhao  and Yuanhua Jia


Research Article (18 pages), Article ID 6610882, Volume 2021 (2021)

## **Museum Display Showcase Furniture System Research Based on Internet of Things Technology in Intelligent Environment**

Jiaojiao Hu, Zhihui Wu , and Lei Jin


Research Article (14 pages), Article ID 6689744, Volume 2021 (2021)

## **Application of Remote Sensing Image Data Scene Generation Method in Smart City**

Yuanjin Xu 


Research Article (13 pages), Article ID 6653841, Volume 2021 (2021)

**Modeling and Research on Human Capital Accumulation Complex System of High-Tech Enterprises Based on Big Data**

Yanan Shen 


Research Article (14 pages), Article ID 6635228, Volume 2021 (2021)

**Development and Supervision of Robo-Advisors under Digital Financial Inclusion in Complex Systems**

Wensheng Dai 


Research Article (12 pages), Article ID 6666089, Volume 2021 (2021)

**Detection and Adaptive Video Processing of Hyperopia Scene in Sports Video**

Qingjie Chen and Minkai Dong 


Research Article (13 pages), Article ID 6610760, Volume 2021 (2021)

**Study on Mechanical Characteristics of Energy-Absorbing and Anti-Scour Bolts**

Zhi Tang , Hao Wu, Jinguo Lv, Zhuangzhuang Xin, and Wenbo Zuo


Research Article (9 pages), Article ID 8876517, Volume 2021 (2021)

**Electroencephalogram of Happy Emotional Cognition Based on Complex System of Music and Image Visual and Auditory**

Lin Gan, Mu Zhang , Jijia Jiang, and Fajie Duan



Research Article (14 pages), Article ID 6631400, Volume 2020 (2020)

**Collaborative Intelligent Environment Perception and Mission Control of Scientific Researchers in Semantic Knowledge Framework Based on Complex Theory**

Jingfeng Zhao and Yan Li 


Research Article (11 pages), Article ID 6637375, Volume 2020 (2020)

**Impact of Financial R&D Resource Allocation Efficiency Based on VR Technology and Machine Learning in Complex Systems on Total Factor Productivity**

Hui Sun  and Xiong Zhong 


Research Article (15 pages), Article ID 6679846, Volume 2020 (2020)

**Simulation Study on Fire Visibility of Typical Floor Planes of Modern Super High-Rise Office Buildings in China**

Tongtong Zhang 

Research Article (14 pages), Article ID 8868522, Volume 2020 (2020)

**Simulation Study on the Spatiotemporal Difference of Complex Neurodynamics between P3a and P3b**

Xin Wei, Xiaoli Ni , Junye Liu, Haiyang Lang, Rui Zhao, Tian Dai, Wei Qin, Wei Jia, and Peng Fang

Research Article (11 pages), Article ID 2796809, Volume 2020 (2020)

**Intelligent Digital Currency and Dynamic Coding Service System Based on Internet of Things Technology**

Shanshen Li  and Xin Jing 

Research Article (16 pages), Article ID 6647039, Volume 2020 (2020)




# Contents

## **Research on Lung Nodule Detection Based on Improved Target Detection Network**

Ye Li , Qian Wu , Hongwei Sun , and Xuewei Wang 




Research Article (7 pages), Article ID 6633242, Volume 2020 (2020)

## **Robust Adaptive Control for a Class of T-S Fuzzy Nonlinear Systems with Discontinuous Multiple Uncertainties and Abruptly Changing Actuator Faults**

Xin Ning, Yao Zhang, and Zheng Wang 


Research Article (16 pages), Article ID 8863073, Volume 2020 (2020)

## **M&A Short-Term Performance Based on Elman Neural Network Model: Evidence from 2006 to 2019 in China**

Ming Xiao , Xionghui Yang , and Ge Li 



Research Article (15 pages), Article ID 8811273, Volume 2020 (2020)

## **The Organizational Structure and Operational Logic of an Urban Smart Governance Information Platform: Discussion on the Background of Urban Governance Transformation in China**

Junfang Kang and Xianjun Wang 

Research Article (16 pages), Article ID 6638958, Volume 2020 (2020)

## **Research on the Impact of Green Finance and Fintech in Smart City**

Zheng He, Zhengkai Liu, Hui Wu , Xiaomin Gu , Yuanjun Zhao, and Xiaoguang Yue


Research Article (10 pages), Article ID 6673386, Volume 2020 (2020)

## **Application Research of Intelligent Classification Technology in Enterprise Data Classification and Gradation System**

Lina Yu , Chunwei Wang , Huixian Chang , Sheng Shen , Fang Hou , and Yingwei Li 

Research Article (9 pages), Article ID 6695484, Volume 2020 (2020)

## **Mechanism of Drag Reduction in Floating Plate of Paddy Field Based on CFD**

Xiaoze Yu, Baofeng Zhang, and Jiahao You 


Research Article (12 pages), Article ID 8852941, Volume 2020 (2020)

## **A New Method for Optimizing the Cabin Layout of Manned Submersibles**

Wenzhong Wang , Shusheng Zhang, Cong Ye, Dengkai Chen , and Hao Fan 


Research Article (11 pages), Article ID 6626602, Volume 2020 (2020)

## **Green Energy Strategic Management for Service of Quality Composition in the Internet of Things Environment**

Jianhao Gao 


Research Article (10 pages), Article ID 6678612, Volume 2020 (2020)

## **Visualization of Information Retrieval in Smart Library Based on Virtual Reality Technology**

Shulin Fang 


Research Article (18 pages), Article ID 6646673, Volume 2020 (2020)

**Landscape Image Layout Optimization Extraction Simulation of 3D Pastoral Complex under Big Data Analysis**

Juan Du  and Yuelin Long

Research Article (11 pages), Article ID 6620216, Volume 2020 (2020)

**Improved Newton Iterative Algorithm for Fractal Art Graphic Design**

Huijuan Chen and Xintao Zheng 


Research Article (11 pages), Article ID 6623049, Volume 2020 (2020)

**Resource Scheduling and Strategic Management of Smart Cities under the Background of Digital Economy**

Qing Yin  and Gang Liu




Research Article (12 pages), Article ID 6624307, Volume 2020 (2020)

**Corrigendum to “Explosion Resistance of Three-Dimensional Mesoscopic Model of Complex Closed-Cell Aluminum Foam Sandwich Structure Based on Random Generation Algorithm”**

Zhen Wang, Wen Bin Gu , Xing Bo Xie, Qi Yuan, Yu Tian Chen, and Tao Jiang

Corrigendum (2 pages), Article ID 9519832, Volume 2020 (2020)

**The Measurement of Chinese Sentence Semantic Complexity**

Shuqin Zhu , Jihua Song , Weiming Peng , Dongdong Guo, and Jingbo Sun


Research Article (10 pages), Article ID 8871263, Volume 2020 (2020)

**Three-Dimensional Finite Element Numerical Simulation and Analysis of Solid-State Processing of Metal Material**

Guang Su and Aimin Zhang 


Research Article (12 pages), Article ID 8819745, Volume 2020 (2020)

**Accurate Recognition and Simulation of 3D Visual Image of Aerobics Movement**

Wenhua Fan and Hyun Joo Min 

Research Article (11 pages), Article ID 8889008, Volume 2020 (2020)

**Moving Vehicle Tracking Optimization Method Based on SPF**

Caixia Lv  and Xuejing Zhang


Research Article (14 pages), Article ID 8829167, Volume 2020 (2020)

**Power Control for Full-Duplex Device-to-Device Underlaid Cellular Networks: A Stackelberg Game Approach**

Zhen Yang, Titi Liu , and Guobin Chen

Research Article (12 pages), Article ID 1786349, Volume 2020 (2020)


**Performance Investigation of Stochastic Resonance in Three Types of Asymmetric Bistable System Driven by Trichotomous Noise**

Si-Hai Zhao , Jiang-Ye Xu, Yu-Xiao Liu, Ze-Xing Zhao, and Zhong-Shun Qin

Research Article (12 pages), Article ID 8842122, Volume 2020 (2020)

# Contents


## **Prediction Mathematic Model and Influencing Factors of Contact Stress of Cylindrical Gear with Arc Tooth**

Qi Zhang , Guang Wen, Shuang Liang, Qin Tong, Li Hou, and Guangchun Yang  
Research Article (15 pages), Article ID 8888407, Volume 2020 (2020)

## **Analysis of Asymmetric Piecewise Linear Stochastic Resonance Signal Processing Model Based on Genetic Algorithm**

Lina He  and Chuan Jiang   
Research Article (11 pages), Article ID 8817814, Volume 2020 (2020)



## **Decision Tree Algorithm-Based Model and Computer Simulation for Evaluating the Effectiveness of Physical Education in Universities**

Zhifei Zhang, Zijian Zhao , and Doo-Seoung Yeom  
Research Article (11 pages), Article ID 8868793, Volume 2020 (2020)



## **Real-Time Visualization Optimization Management Simulation of Big Data Stream on Industrial Heritage Cloud Platform**

Mengya Gao   
Research Article (10 pages), Article ID 8885191, Volume 2020 (2020)



## **Application of Empirical Orthogonal Function Interpolation to Reconstruct Hourly Fine Particulate Matter Concentration Data in Tianjin, China**

Hongwu Zhou, Haidong Pan, Shuang Li , and Xianqing Lv   
Research Article (15 pages), Article ID 9724367, Volume 2020 (2020)






## **Research on the Measurement, Evolution, and Driving Factors of Green Innovation Efficiency in Yangtze River Economic Belt: A Super-SBM and Spatial Durbin Model**

Renyang Long , Hangyuan Guo, Danting Zheng, Ronghua Chang, and Sanggyun Na   
Research Article (14 pages), Article ID 8094247, Volume 2020 (2020)



## **Optimization and Simulation of Controller Area Network Communication Model Based on Industrial Internet of Things Platform**

Haifeng Lin  and Lin Du   
Research Article (11 pages), Article ID 8864722, Volume 2020 (2020)

## **Heuristic Sensing: An Uncertainty Exploration Method in Imperfect Information Games**

Zhenyang Guo , Xuan Wang , Shuhan Qi , Tao Qian , and Jiajia Zhang   
Research Article (9 pages), Article ID 8815770, Volume 2020 (2020)

## **A Dynamic Hierarchical Clustering Data Gathering Algorithm Based on Multiple Criteria Decision Making for 3D Underwater Sensor Networks**


Xiaoying Song , Wei Sun, and Qilong Zhang   
Research Article (14 pages), Article ID 8835103, Volume 2020 (2020)

**Hemodynamic-Based Evaluation on Thrombosis Risk of Fusiform Coronary Artery Aneurysms Using Computational Fluid Dynamic Simulation Method**

Haoran Wang, Hitomi Anzai, Youjun Liu, Aike Qiao , Jinsheng Xie, and Makoto Ohta 


Research Article (11 pages), Article ID 8507273, Volume 2020 (2020)

**Optimization and Simulation for E-Commerce Supply Chain in the Internet of Things Environment**

Liwei Zhu 


Research Article (11 pages), Article ID 8821128, Volume 2020 (2020)

**A Three-Dimensional Complex Measurement Model-Based Avionic Radio-Frequency Power Source Health Assessment Method**

Lin Huo , Shiqi Li, Simiao Fei, and Chuan Lyu


Research Article (14 pages), Article ID 6369729, Volume 2020 (2020)

**Multiple-Devices-Process Integrated Scheduling Algorithm with Time-Selective Strategy for Process Sequence**

Xiaohuan Zhang , Dan Zhang, Zhen Wang, and Yu Xin


Research Article (12 pages), Article ID 8898536, Volume 2020 (2020)

**Panoramic Display and Planning Simulation of Civil Engineering Project Based on Virtual Reality Technology**

Zheng Wang , Chang Liu, and Bin Song



Research Article (11 pages), Article ID 1424757, Volume 2020 (2020)

**Research of the Context Recommendation Algorithm Based on the Tripartite Graph Model in Complex Systems**

Fei Long 


Research Article (11 pages), Article ID 7945417, Volume 2020 (2020)

**Numerical Investigation on the Dynamic Characteristics of an Adjustable Power Turbine Used in Environmental Control System**

Qihang Lu , Qi Zhang, and Dalin Zhang 

Research Article (12 pages), Article ID 7916936, Volume 2020 (2020)

**Diagnosis and Exercise Rehabilitation of Knee Joint Anterior Cruciate Ligament Injury Based on 3D-CT Reconstruction**

Shunchao Zhang 

Research Article (13 pages), Article ID 3690124, Volume 2020 (2020)


**Big Data Analytics for Complex Credit Risk Assessment of Network Lending Based on SMOTE Algorithm**

Aiwen Niu, Bingqing Cai, and Shousong Cai 

Research Article (9 pages), Article ID 8563030, Volume 2020 (2020)


# Contents

## **Articulatory-to-Acoustic Conversion Using BiLSTM-CNN Word-Attention-Based Method**

Guofeng Ren , Guicheng Shao, and Jianmei Fu

Research Article (10 pages), Article ID 4356981, Volume 2020 (2020)

## **Optimization and Simulation of a Reasonable Scheduling Model under Multiple Tasks in Company Management**

Zhuo Wang 


Research Article (10 pages), Article ID 1204815, Volume 2020 (2020)

## **A 3D Reconstruction Method Using Multisensor Fusion in Large-Scale Indoor Scenes**

Panlong Gu , Fengyu Zhou , Dianguo Yu , Fang Wan , Wei Wang , and Bangguo Yu 



Research Article (14 pages), Article ID 6973790, Volume 2020 (2020)

## **Meso-Complexity Computer Simulation Investigation on Antiexplosion Performance of Double-Layer Foam Aluminum under Pore Grading**

Zhen Wang, Wen Bin Gu , Xing Bo Xie, Yu Tian Chen, and Lei Fu

Research Article (13 pages), Article ID 4121926, Volume 2020 (2020)

## **Dynamic Response of a Semiactive Suspension System with Hysteretic Nonlinear Energy Sink Based on Random Excitation by means of Computer Simulation**

Hui Chen  and Wuyin Jin 

Research Article (13 pages), Article ID 3181423, Volume 2020 (2020)

## **A Deep Learning Approach for a Source Code Detection Model Using Self-Attention**

Yao Meng  and Long Liu


Research Article (15 pages), Article ID 5027198, Volume 2020 (2020)

## **Photovoltaic Generation Prediction of CCIPCA Combined with LSTM**

E. Zhu  and D. Pi

Research Article (11 pages), Article ID 1929372, Volume 2020 (2020)

## **Analysis and Simulation of Multimedia English Auxiliary Handle Based on Decision Tree Algorithm**

Kaiwei Yan 


Research Article (12 pages), Article ID 9028303, Volume 2020 (2020)

## **Emotion Recognition of Students Based on Facial Expressions in Online Education Based on the Perspective of Computer Simulation**

Weiqing Wang, Kunliang Xu, Hongli Niu, and Xiangrong Miao 





Research Article (9 pages), Article ID 4065207, Volume 2020 (2020)

## **An Innovative Aerial Manipulator with Tandem Ducted Fans: Modeling, Control, and Simulation**

Yibo Zhang, Wei Fan , Changle Xiang, Bin Xu, Tianfu Ai, Lei Yuan, and Yang Liu

Research Article (24 pages), Article ID 7923539, Volume 2020 (2020)

**Research on the Niche Evolution Game of Ecological Community Innovation of Corporate Venture Capital Based on Logistic Extended Complexity Model**

Fanglin Meng , Zengrui Tian , Beiquan Chang , Hongxin Yu , and Shuai Zhang  
Research Article (13 pages), Article ID 6327218, Volume 2020 (2020)

**Multiobject Detection Algorithm Based on Adaptive Default Box Mechanism**

Jinling Li , Qingshan Hou , and Jinsheng Xing   
Review Article (11 pages), Article ID 5763476, Volume 2020 (2020)


**Hybrid Depth-Separable Residual Networks for Hyperspectral Image Classification**

Cuijie Zhao, Hongdong Zhao , Guozhen Wang, and Hong Chen  
Research Article (17 pages), Article ID 4608647, Volume 2020 (2020)


**Modeling and Simulation of English Speech Rationality Optimization Recognition Based on Improved Particle Filter Algorithm**

Hui Dong   
Research Article (10 pages), Article ID 6053129, Volume 2020 (2020)


**Numerical Simulation of Ambiguity Resolution in Multiple Information Streams Based on Network Machine Translation**

Lei Wang and Qun Ai   
Research Article (10 pages), Article ID 7278085, Volume 2020 (2020)


**Emotional Interactive Simulation System of English Speech Recognition in Virtual Context**

Dan Li   
Research Article (11 pages), Article ID 9409630, Volume 2020 (2020)

**Explosion Resistance of Three-Dimensional Mesoscopic Model of Complex Closed-Cell Aluminum Foam Sandwich Structure Based on Random Generation Algorithm**

Zhen Wang, Wen Bin Gu , Xing Bo Xie, Qi Yuan, Yu Tian Chen, and Tao Jiang  
Research Article (16 pages), Article ID 8390798, Volume 2020 (2020)


**Study on Global Parameters Optimization of Dual-Drive Powertrain System of Pure Electric Vehicle Based on Multiple Condition Computer Simulation**

Yong Wang, Hongguo Cai , Yinghua Liao, and Jun Gao  
Research Article (10 pages), Article ID 6057870, Volume 2020 (2020)

**Optimal Pricing Decision of Fashion Apparel considering Experiential Service and Design Enhancement**

Qi Xu , Lili Zhou , and Qi Chen   
Research Article (11 pages), Article ID 9505120, Volume 2020 (2020)

**Application Study of Sigmoid Regularization Method in Coke Quality Prediction**

Shaohong Yan, Hailong Zhao, Liangxu Liu, Qiaozhi Sang, Peng Chen, and Jie Li   
Research Article (10 pages), Article ID 8785047, Volume 2020 (2020)

# Contents

## **A Research on the Control System of High-Speed Homopolar Motor with Solid Rotor Based on Flywheel Energy Storage**

Lili Jing, Yandong Yu , and Xiaochuan Xue


Research Article (12 pages), Article ID 6537563, Volume 2020 (2020)

## **DTFNet: Dynamic and Texture Features Fusion Attention Network for Face Antispoofing**

Xin Cheng , Hongfei Wang , Jingmei Zhou , Hui Chang , Xiangmo Zhao , and Yilin Jia


Research Article (11 pages), Article ID 5836596, Volume 2020 (2020)

## **Semiparametric Deep Learning Manipulator Inverse Dynamics Modeling Method for Smart City and Industrial Applications**

Nan Liu, Liangyu Li, Bing Hao, Liusong Yang, Tonghai Hu, Tao Xue, Shoujun Wang , and Xingmao Shao


Research Article (11 pages), Article ID 9053715, Volume 2020 (2020)

## **An Anonymous Authentication Scheme in VANETs of Smart City Based on Certificateless Group Signature**

Yuanpan Zheng , Guangyu Chen, and Liguan Guo





Research Article (7 pages), Article ID 1378202, Volume 2020 (2020)

## **Adaptive Language Processing Based on Deep Learning in Cloud Computing Platform**

Wenbin Xu  and Chengbo Yin

Research Article (11 pages), Article ID 5828130, Volume 2020 (2020)

## **A Multiscale and High-Precision LSTM-GASVR Short-Term Traffic Flow Prediction Model**

Jingmei Zhou , Hui Chang , Xin Cheng , and Xiangmo Zhao 

Research Article (17 pages), Article ID 1434080, Volume 2020 (2020)

## **Research on Complex Classification Algorithm of Breast Cancer Chip Based on SVM-RFE Gene Feature Screening**

Guobin Chen, Xianzhong Xie, and Shijin Li 

Research Article (12 pages), Article ID 1342874, Volume 2020 (2020)

## **Automatic Grading for Complex Multifile Programs**

Tiantian Wang , Djoko Budi Santoso, Kechao Wang, and Xiaohong Su

Research Article (15 pages), Article ID 3279053, Volume 2020 (2020)

## **A Hemodynamic-Based Evaluation of Applying Different Types of Coronary Artery Bypass Grafts to Coronary Artery Aneurysms**

Haoran Wang, Hitomi Anzai, Youjun Liu, Aike Qiao , Jinsheng Xie, and Makoto Ohta 

Research Article (10 pages), Article ID 9359340, Volume 2020 (2020)

## **Study on the 2D Optimization Simulation of Complex Five-Hole Cutting Blasting under Different Lateral Pressure Coefficients**

Jing Gao, Shizhen Xie, Xiantang Zhang , Hongli Wang, Wenle Gao, and Hongmin Zhou


Research Article (12 pages), Article ID 4639518, Volume 2020 (2020)

**A Bayesian Network under Strict Chain Model for Computing Flow Risks in Smart City**

Zengfanxiang Wei , Lei Zhang , Qi Yue , and Muchen Zhong 


Research Article (8 pages), Article ID 5920827, Volume 2020 (2020)

**Research on Sentiment Tendency and Evolution of Public Opinions in Social Networks of Smart City**

Yanni Liu, Dongsheng Liu , and Yuwei Chen







Research Article (13 pages), Article ID 9789431, Volume 2020 (2020)

**A Complexity Analysis of User Interaction with Hotel Robots**

Lina Zhong , Liyu Yang, Jia Rong, and Xiaonan Li

Research Article (13 pages), Article ID 4537152, Volume 2020 (2020)

**Decision Optimization of Low-Carbon Dual-Channel Supply Chain of Auto Parts Based on Smart City Architecture**

Zheng Liu , Bin Hu , Bangtong Huang , Lingling Lang , Hangxin Guo , and Yuanjun Zhao 

Research Article (14 pages), Article ID 2145951, Volume 2020 (2020)



## Editorial

# Complexity Problems Handled by Advanced Computer Simulation Technology in Smart Cities 2021

Zhihan Lv <sup>1</sup>, Kaoru Ota,<sup>2</sup> Jaime Lloret <sup>3</sup>, Wei Xiang,<sup>4</sup> and Paolo Bellavista <sup>5</sup>

<sup>1</sup>Qingdao University, Qingdao, China

<sup>2</sup>Muroran Institute of Technology, Muroran, Japan

<sup>3</sup>Instituto de Investigacion para la Gestion Integrada de Zonas Costeras, Universitat Politecnica de Valencia, Spain

<sup>4</sup>James Cook University, Cairns, Australia

<sup>5</sup>Università di Bologna, Bologna, Italy

Correspondence should be addressed to Zhihan Lv; lvzhihan@gmail.com

Received 5 January 2022; Accepted 5 January 2022; Published 2 February 2022

Copyright © 2022 Zhihan Lv et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Complexity science technology will bring opportunities for the continuous expansion of new simulation research fields, especially for human simulation, social simulation, and human brain simulation, and provide unprecedented opportunities for simulation for social governance, prediction, and urbanization. The development of complexity science technology provides an important opportunity to use modeling and simulation methods to study new virtual information spaces. As we all know, the advantage of modeling and simulation technology is to provide a bridge from the real world to the virtual space, which provides a substitute for studying the unknown fields of human society and the physical world.

However, virtual information spaces are a completely new virtual space that are completely different from the physical world and human society. This virtual space is not only “human, machine, and object” but also a space where the three notions are combined, and there are also the characteristics of multilayer mesh and cross-domain association. The study of this new virtual space is a complex and current scientific research area. On the basis of making full use of the results of complexity science research, using parallel systems, embedded simulation, and other methods to build a bridge to this new type of space, they provide a virtual alternative for the interaction or behavioral characteristics between the research. Using the results of complexity science to constantly modify and test the models built is an issue that needs to be addressed for the further development of modeling and simulation science.

This special issue collates 189 original research papers and comments with a focus on up-to-date technology related to computer simulation for complexity issues in smart cities and serves as a forum for researchers all over the world to discuss their work and recent advances in this field. In particular, this special issue aims to showcase the most recent achievements and developments in complexity problem discovery and exploration.

With the development of science and technology, the demand of human-computer interaction has increased, and the research on human-computer interaction in real and virtual scenarios has become more and more extensive. The use of virtual reality technology and the combination of human-computer interaction technology have been increasingly used in the fields of simulated sports, medical rehabilitation, and game creation. Action is the basis of human behaviours, among which human behaviours and movement analysis are an important research direction. Among human behaviours and actions, the recognition research based on behaviours and actions has the characteristics of convenience, intuition, strong interaction, and rich expression information. It has become the first choice for many researchers to analyse human behaviours. At the same time, the study of some complex objects such as human movement has many fuzzy factors, which are difficult to express and deal with. A total of 26 research papers in this special issue propose processing methods and optimization algorithms for human-computer interaction applications in multiple scenarios, providing solutions to complex human-

computer interaction problems from many aspects, and all have good application performance.

Urban digital twins are virtual representations of urban physical assets and are an important development direction of digital twin technology. Smart cities built under the support of digital twin technology can bring better cost efficiencies, operational efficiencies, better crisis management, more openness and better-informed decision-making, more participatory governance, or better urban planning. The special issue contains 5 representative research papers on smart city construction. I. Meta et al. proposed a modular framework for urban digital twins, combined with the concept of urban physiology, to better solve the theoretical problems of smart city construction. Y. He et al. proposed a knowledge graph fusion framework for building smart cities. S. Wang et al. provided a feature extraction method for the spatial distribution of urban landforms. Starting from the energy efficiency management of smart cities, L. Chen and P. Han constructed an efficient system platform based on the mobile data of the Internet of Things. H. Wang et al. improved the KNN algorithm and optimized the preprocessing process of the city center. In addition, the special issue also includes a total of 129 papers on the optimization methods of urban life in other fields, which include research methods from different perspectives of urban development and urban life, which optimize and enrich urban construction and daily need to a great extent. These areas include education, healthcare, construction, tourism development, financial transactions, entertainment consumption, urban security services, low-carbon and green development, urban resource allocation and control optimization, environmental and geological monitoring, urban production, sports, and urban transportation. In these areas, this special issue brings together a wealth of optimization methods.

In the field of education, the complexity of the education system at all levels has gradually deepened. Whether it is a junior high school, high school, or university, excellent student management systems, online education systems, teaching quality evaluation, and more intelligent teaching models are the main development directions. The special issue contains 32 education-related papers. Starting from the practical development of subject education in many aspects, a series of optimization algorithms and platform construction methods are proposed. In addition, this is not only limited to conventional teaching types but also includes painting, sports, and student psychological supervision, which promote the intelligent development of education in a more comprehensive manner.

In terms of healthcare, Y. Chen proposed an optimized plan for the clinical nursing management system to make the overall clinical nursing process more perfect. Y. Jiang et al. also proposed an optimized method for analyzing clinical information. Aiming at the fusion of supporting medical image data on the ground, R. Chen et al. proposed a system framework that breaks through some of the limitations of traditional image fusion. X. Che compared the difference between exercise intervention and other nondrug therapies. B. Yang and S. Wang investigated the cognitive function.

For the field of urban building structure, here are some further improvements based on typical treatment methods. J. Yu et al. put forward a method on how to achieve healthy development of contaminated areas and facility construction. M. Wei combined augmented reality to realize more intuitive visualization of building data and intelligent detection. Other papers include such as the treatment method of building fire protection, the distribution method of housing construction in some newly developed areas, the virtual reality construction technology of landscape architecture, the intelligent combination of architecture and environment, and landscape space layout or landscape planning and design method. The special issue contains 4 papers on tourism. With the development of the national economy and society, the scale of the tourism market has expanded rapidly. The informatization and intelligent construction of scenic spots cannot keep up with the pace of economic development, and there is a lack of effective management methods to predict or even improve the overloading of scenic spots, which objectively leads to problems of congestion and overload. More and more tourists flock to the scenic area, which brings greater pressure and safety hazards to the scenic area. The relevant research in this part effectively responded to this problem, formulated a suitable flow control plan, and promoted the intelligent development of scenic spot management.

In the field of urban economy, it includes financial transactions, entertainment consumption, and urban security services. The problem of economic structure equilibrium and its development trend have far-reaching theoretical significance and important practical value. In theory, the study of economic structure equilibrium is an intuitive reflection and detailed elaboration of the status quo of economic development and is an extension and expansion of economic research theories. The changes in the economic structure are more complicated, and the problem of building intelligent economic industry forecasting models needs to be solved urgently. It is very important to study the balance of economic structure and economic development trends and to carry out macrocontrol on the status and trends of economic development. This part contains 39 research papers and 1 comment. In terms of entertainment consumption, the special issue includes movie recommendation, movie big data analysis, and other movie-related applications. At the same time, there are many papers about music feature optimization, music signal segmentation retrieval, and other applications. In urban security, social security services, logistics services, and effective optimization of news reports promote the lives of residents and social security from multiple perspectives. By combining advanced technologies such as big data, Internet of Things, and artificial intelligence, it has promoted the intelligent development of urban economy and entertainment guarantee.

A total of 20 papers were included to discuss the low-carbon and green development of the city, the allocation of urban resources, and the optimization of control. Against the background of the development of innovative countries, climate change and environmental issues have become increasingly prominent. Low-carbon economy has become an important choice for the future economic development of all

countries, and the green development of cities is an inevitable trend. Representatives such as the twin cities proposed by Y. Liu et al. provide a new perspective for the development of urban intelligence. H. Liu integrated low-carbon concepts into urban development. Xu et al. proposed digital technology to help food supply and provided corresponding optimization solutions. L. Zhang built a marketing management system based on cloud computing and big data. S. Mao and R. Huang also proposed intelligent construction methods for marketing strategies. The development of emerging technologies improves the basic appearance of cities by regulating the rational allocation of urban resources. Urban services also include research on sports, urban production, basic transportation, and environmental monitoring. Among them, there are 9 sports-related papers, and most of the papers contain related concepts of human-computer interaction, which can realize the monitoring and modeling of complex movements in some sports, and make a certain contribution to the intelligent development of urban sports. In terms of environmental monitoring, J. Jin et al. provided supporting technologies for geological monitoring in important strategic areas in some countries.

Some new data processing methods in the background of big data are also proposed here. X. Wang et al. proposed a digital image processing method based on the geographic information system. X. Xu et al. combined virtual reality technology to realize the restoration of similar images. S. Wu et al. used virtual reconstruction technology to acquire 3D images. X. Liu created a 3D modeling method that is more in line with image sequences. In addition, the special issue includes 21 algorithm optimization methods, covering most fields such as deep learning, blockchain distributed computing, edge computing, animation synthesis, and semantic similarity algorithms, providing many algorithm optimizations for different field methods.

This special issue has made important contributions to all the fields mentioned above and has promoted the development of the complexity science of smart cities from multiple perspectives. Combining the new generation of information technology and big data technology in the Innovation 2.0 environment, it strengthens open data and social participation in smart cities, solves the problems of open data and privacy protection in the context of big data, and lays the foundation for the construction of new smart cities.

### **Conflicts of Interest**

The editors declare that there are no conflicts of interest regarding the publication of this paper.

*Zhihan Lv  
Kaoru Ota  
Jaime Lloret  
Wei Xiang  
Paolo Bellavista*

## Research Article

# Analysis of Feature Extraction and Anti-Interference of Face Image under Deep Reconstruction Network Algorithm

Jin Yang , Yuxuan Zhao , Shihao Yang , Xinxin Kang , Xinyan Cao, and Xixin Cao

College School of Software & Microelectronics, Peking University, Beijing 102600, China

Correspondence should be addressed to Jin Yang; yang.jin@pku.edu.cn

Received 19 July 2020; Revised 11 September 2020; Accepted 23 September 2020; Published 20 March 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Jin Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In face recognition systems, highly robust facial feature representation and good classification algorithm performance can affect the effect of face recognition under unrestricted conditions. To explore the anti-interference performance of convolutional neural network (CNN) reconstructed by deep learning (DL) framework in face image feature extraction (FE) and recognition, in the paper, first, the inception structure in the GoogleNet network and the residual error in the ResNet network structure are combined to construct a new deep reconstruction network algorithm, with the random gradient descent (SGD) and triplet loss functions as the model optimizer and classifier, respectively, and it is applied to the face recognition in Labeled Faces in the Wild (LFW) face database. Then, the portrait pyramid segmentation and local feature point segmentation are applied to extract the features of face images, and the matching of face feature points is achieved using Euclidean distance and joint Bayesian method. Finally, Matlab software is used to simulate the algorithm proposed in this paper and compare it with other algorithms. The results show that the proposed algorithm has the best face recognition effect when the learning rate is 0.0004, the attenuation coefficient is 0.0001, the training method is SGD, and dropout is 0.1 (accuracy: 99.03%, loss: 0.0047, training time: 352 s, and overfitting rate: 1.006), and the algorithm proposed in this paper has the largest mean average precision compared to other CNN algorithms. The correct rate of face feature matching of the algorithm proposed in this paper is 84.72%, which is higher than LetNet-5, VGG-16, and VGG-19 algorithms, the correct rates of which are 6.94%, 2.5%, and 1.11%, respectively, but lower than GoogleNet, AlexNet, and ResNet algorithms. At the same time, the algorithm proposed in this paper has a faster matching time (206.44 s) and a higher correct matching rate (88.75%) than the joint Bayesian method, indicating that the deep reconstruction network algorithm proposed in this paper can be used in face image recognition, FE, and matching, and it has strong anti-interference.

## 1. Introduction

FR technology has been extensively adopted in identity recognition, but it is mainly used to detect biological features in the face for recognition, with strong uniqueness and security [1]. However, the dataset made up of face images is one that presents a highly nonlinear distribution, and if simple classification method is applied, higher classification errors will occur due to individual differences [2]. Moreover, the face image will be disturbed by lighting, decorations, scenes, and other factors during the shooting, which makes FR extremely difficult. At present, the methods commonly used for face detection include principal component analysis, support vector machine, CNN, and active deformation model [3].

Depth model has been applied in many fields, and its application in image recognition is the first to be concerned. Moreover, the depth model is used for FE in the image, which is far better than manual FE, and can be effectively applied in the field where manual FE is not perfect [4]. The common feature of DL and most machine learning is the ability to extract features. In DL, multilayer network structure is mostly applied, which can fuse the bottom features in the image to form the top features [5]. ImageNet's annual ImageNet Large-Scale Visual Recognition Challenge is the largest and most advanced image recognition competition in the world. Since 2012, AlexNET used DCNN to reduce the recognition error rate of the top 5 to 16.4%, and DCNN algorithm has been adopted in the subsequent champion recognition models [6, 7].

The rest of the paper is organized as follows. Section 2 discusses the related work of the DCNN network and image FE technology. Then, in Section 3, a deep reconstruction network is proposed, and the face recognition and facial FE algorithm is constructed based on the optimized deep reconstruction network. Section 4 carries out the verification of face recognition and FE and matching algorithms proposed in this paper and compares them with other DCNN algorithms. Section 5 discusses and analyzes the results obtained in this paper, and the results of the paper were compared with other people's relevant research results. Section 6 concludes the paper with summary and points out the future research directions.

## 2. Related Work

NN is a common computer model, which is mainly based on the traditional statistical modeling of complex relationships to explore the relationship between different datasets. The NN structure in different application fields is different, but the NN structure with the largest application in image-related fields is CNN. Seeliger et al. applied DCNN to the detection of magnetoencephalogram signal map and reconstructed the cerebral cortex activity model based on the extracted resolution source [8]. Hoang and Kang proposed a method for fault diagnosis of rolling axis based on CNN and vibration images based on the depth structure and found that this method does not need FE technology and has high diagnostic accuracy and robustness under noisy environment [9]. At present, deep learning has been widely used in the field of image analysis, especially in facial image recognition, which has obtained excellent results. Zhang et al. proposed a convolutional neural network model based on three-stage filtering and HOG, which was applied to the recognition of color face images and obtained excellent effect [10]. Goswami et al. proposed a framework for face recognition based on deep neural networks and used a characteristic anomaly filter to detect singularities. Finally, after verification with public data, it was found that the model built by it had strong face recognition robustness [11]. Isogawa et al. indicated that there was no denoising method capable of parameter adjustment in the DCNN model.

They used the proportional coefficient of soft shrinkage threshold to optimize DCNN, which was found to be applicable to noise removal in the image [12]. Nakazawa and Kulkarni proposed a method to detect and segment abnormal defect patterns using the NN architecture of deep convolutional encoder-decoder and finally found that these models could detect invisible defect patterns from real images [13].

The basic principle of image FE is to use a computer to extract the information in the image and then judge whether the differences in the image are one of the features in the image. An and Zhou proposed an image FE algorithm based on two-dimensional empirical mode decomposition and scale-invariant feature change. After verification, they found that it could effectively improve the speed and accuracy of FE [14]. Xu et al. proposed a model for automatically extracting noise level perception features based on the CNN model and

feature vectors and found that the proposed algorithm had high extraction speed and accuracy at different noise levels [15]. Fieldin et al. used enhanced adaptive particle swarm optimization to conduct deep CNN evolution, and after training and verification, they found that the extraction error rate was 4.78% [16].

To sum up, it is evident that DCNN is widely used in image recognition. The improved DCNN can effectively extract the features in the image. However, there are few research studies on the anti-interference of DCNN in face image recognition. Therefore, a DCNN based on Caffe depth framework is proposed. Matlab simulation software is used to explore the influence of different parameter settings on the performance of FR by constructing the DCNN model. Then, the LFW database and the self-constructed face database in this paper are combined to explore the anti-interference performance of the DCNN model constructed in this paper for face image recognition in different scenarios. The results of this paper are intended to lay a foundation for improving the efficiency of FR.

## 3. Methodology

*3.1. Design of Deep Reconstruction Network Structure.* Existing studies have shown that, with the gradual deepening of the CNN structure, the CNN training results become better, but while improving the results, it will also increase the amount of network computation [17]. Therefore, an inception network structure is proposed in the GoogleNet network, which can make full use of the features extracted by each layer of the network, and the structure can increase the depth and complexity of the network while ensuring the network computing complexity. In the paper, some network parameter adjustments are proposed for the basic structure of the GoogleNet network: (1) the size of the feature map of the input network should be slowly reduced in the network to avoid the bottleneck of the feature representation in the image; (2) high-dimensional features are easier to obtain than low-dimensional features and can accelerate the training speed of the model; (3) the adjacent neurons in the network have close correlations and can be integrated into spatial relations in low-dimensional spaces.

In the convolutional neural network, the  $5 * 5$  size convolution kernel is  $25/9$  times the size of  $3 * 3$  convolution. If the larger convolution kernel can be replaced with the superposition of multiple  $3 * 3$  size convolution kernels, then it can make the network have fewer parameters in the same field of view. Therefore, a hypothesis is proposed in the study. First, a  $3 * 3$  size convolution filter is adopted to process a  $5 * 5$  size block to obtain a  $3 * 3$  size output feature map. After the convolution with a  $3 * 3$  size convolution kernel,  $1 * 1$  size output is obtained. At this time, the final calculation amount is about  $18/25$  of the direct use of the  $5 * 5$  size convolution kernel, which saves 28% of the calculation amount, but the network also needs to use the ReLu activation function to correct the output. After that, the  $2 * 2$  size convolution kernel is adopted to replace the  $3 * 3$  size convolution kernel, which can save 11% of the calculation, but the use of an asymmetric convolution kernel

is better. Therefore, in the study,  $1 * 3$  and  $3 * 1$  size cascaded convolution kernels are used to replace  $3 * 3$  size convolution kernels to save 33% of the calculation, so the larger size convolution kernel can be decomposed in the same way. Based on this, an improved Inception structure is obtained, as shown in Figure 1.

The structure of ResNet can have up to 152 layers [18]. This model adds an identical high-speed channel to a simple stacked shallow network, as shown in Figure 2. On the left is the ordinary network, and its output value is  $H(x)$ , which represents any ideal feature transformation. The residual network output value on the right is  $H(x) = F(x) + x$ , and what needs to be fitted in the network is the residual  $F(x)$ . Adding an identical shortcut connection in the network Europe can make the input mapping and input of the network overlap. The output formula of the residual unit is  $y = F(x, \{Wi\}) + x$ , where  $x$  represents the input and  $y$  represents the output, and the function  $F(x, \{Wi\})$  is the residual mapping obtained by learning. Theoretically, the following two networks can approximate any function. But if the optimal solution of the network is close to the identity mapping, it is easier to optimize using the residual network; if the convolution branch appears as gradient dispersion, the features and gradients in the network can also flow in the branches of the identity mapping, which ensures the reliability of information transmission.

The convolution layer in the convolution branch  $F(x, \{Wi\})$  of the residual unit can be composed of multiple layers, and the residual unit can also be modified by scaling of the mapping branch, convolution kernel size, dropout, and the activation function location. Therefore, a new DCNN network structure is constructed based on the inception structure in the GoogleNet network and the residual structure in the ResNet network, to improve the network's computational efficiency and information circulation effect. The combined new network structure is shown in Figure 3.

The basic parameters of the deep reconstruction network constructed in this paper are shown in Table 1, where the dropout parameter is  $x$ , and the subsequent continuous adjustment of the parameters is required to select the optimal dropout value for the final test.

Deep neural networks often use Softmax as a classifier, and the triplet loss function is a classifier function proposed by Google for face recognition CNNs. Studies have also shown that its recognition rate on the LFW dataset has reached 99.63%, and the recognition rate on the YouTube dataset has also reached 95.12% [19]. Therefore, the impact of different classifiers on the model's face recognition rate is analyzed. In this paper, triplets are composed of a reference (anchor) of random samples in the training set, a reference sample (positive), and a sample of different categories (negative). Assuming that the sample features obtained by extracting each sample in triplets are  $f(x_1)$ ,  $f(x_2)$ , and  $f(x_3)$ , respectively, then the goal of optimization is as follows:

$$\|f(x_1) - f(x_2)\|_2^2 + \alpha < \|f(x_1) - f(x_3)\|_2^2, \quad \forall (x_1^i, x_2^i, x_3^i) \in \tau, \quad (1)$$

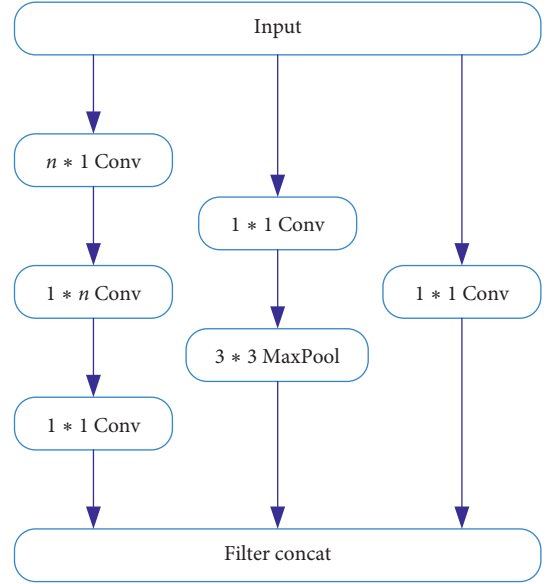


FIGURE 1: Improved inception structure.

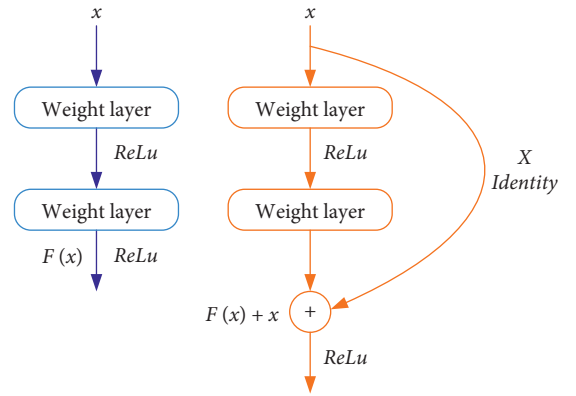


FIGURE 2: Differences in the structure of different networks. The left is an ordinary network, and the right is a residual network in ResNet.

in which  $\alpha$  is the Euclidean distance between sample features of the same category or the minimum interval of Euclidean distance between sample features of different categories.

At this time, the mathematical expression of triplet loss is as follows:

$$L = \sum_i^N \left[ \|f(x_1^i) - f(x_2^i)\|_2^2 - \|f(x_1^i) - f(x_3^i)\|_2^2 + \alpha \right]. \quad (2)$$

In the training process of the neural network, the gradient equation corresponding to each parameter is as follows:

$$\frac{\partial L}{\partial f(x_1^i)} = 2(f(x_3^i) - f(x_2^i)), \quad (3)$$

$$\frac{\partial L}{\partial f(x_2^i)} = 2(f(x_2^i) - f(x_1^i)), \quad (4)$$

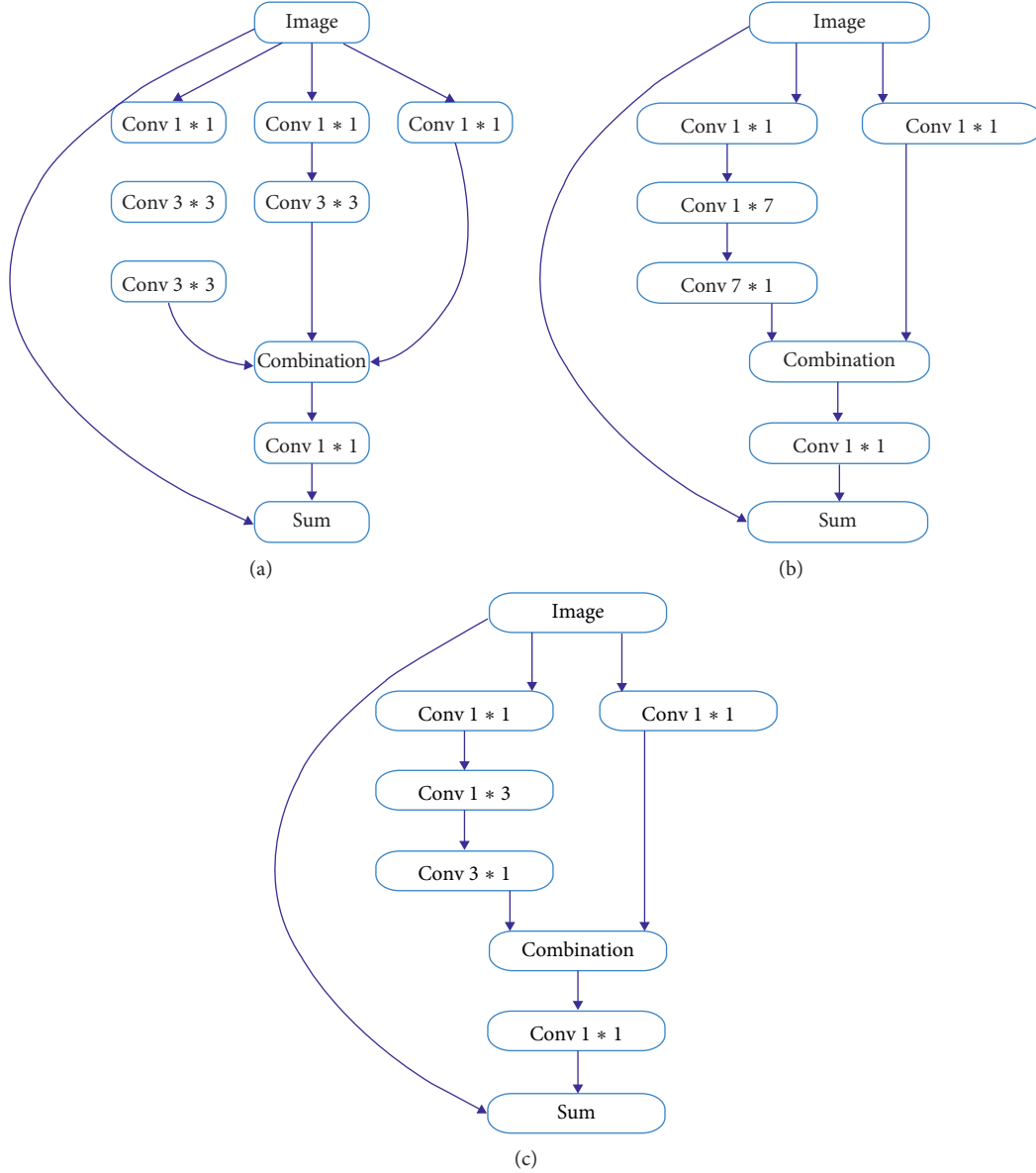


FIGURE 3: (a) ResNet-Inception-1 structure; (b) ResNet-Inception-2 structure; (c) ResNet-Inception-3 structure.

$$\frac{\partial L}{\partial f(x_3^i)} = 2(f(x_1^i) - f(x_3^i)). \quad (5)$$

In order to improve the optimization efficiency of the model, the online method is used to select the combination of triplets. At the same time, in order to prevent the model from falling into the local optimal solution state,  $\alpha$  is removed when selecting the negative sample, so  $x_n^i$  only needs to satisfy

$$\|f(x_1^i) - f(x_3^i)\|_2^2 < \|f(x_1^i) - f(x_n^i)\|_2^2. \quad (6)$$

Center loss can ensure the separation of samples of different categories during the training process and also ensure the aggregation characteristics between samples of

the same category [20]. Center loss can be defined as follows: in which  $c_{yi}$  is the feature center of the  $i$ -th category.

$$L_C = \frac{1}{2} \sum_{i=1}^m \|x_i - c_{yi}\|_2^2, \quad (7)$$

where  $L_C$  is the center loss.  $C_{yi}$  is the feature center of  $i$ -th category.  $x_i$  is the feature before full connection layer, and  $m$  is the size of the mini-batch.

Therefore, the equation expects the smaller the sum of the distances between the characteristics and the feature center of each sample in the batch, that is, the smaller the distance within the class, the better the performance is.

TABLE 1: Basic parameters of ResNet-Inception network structure.

| Structure          | Parameter | Output image size |
|--------------------|-----------|-------------------|
| Conv-1             | 3 * 3     | 147 * 147 * 32    |
| Conv-2             | 3 * 3     | 145 * 145 * 32    |
| Conv-3             | 3 * 3     | 145 * 145 * 64    |
| Pool-1             | —         | 73 * 73 * 64      |
| Conv-4             | 1 * 1     | 73 * 73 * 80      |
| Conv-5             | 3 * 3     | 70 * 70 * 190     |
| Conv-6             | 3 * 3     | 33 * 33 * 256     |
| ResNet-Inception-1 | 5         | 33 * 33 * 256     |
| Reduction-1        | —         | 16 * 16 * 512     |
| ResNet-Inception-2 | 10        | 16 * 16 * 512     |
| Reduction-2        | —         | 8 * 8 * 1024      |
| ResNet-Inception-3 | 5         | 8 * 8 * 1024      |
| Pool-2             | —         | 1792              |
| Dropout            | $x$       | 1792              |

The  $c_{y_i}$  is updated to update the sample set in the input network, and the update rule is as follows:

$$c_u^{v+1} = c_u^v - \frac{\sum_{i=1}^n \delta(y_i = u) \cdot (c_{y_i} - x_i)}{1 + \sum_{i=1}^n \delta(y_i = u)}, \quad (8)$$

in which  $\delta$  belongs to 0 or 1.

Then, the error function in the network can be defined as follows:

$$\text{LOSS} = \text{LOSS}_s + \lambda \text{LOSS}_c, \quad (9)$$

in which  $\lambda$  is the weight in the loss function. When  $\lambda$  is 0, the classifier in the network is Softmax.

**3.2. Face Fe and Matching Based on Deep Reconstruction Network.** In the process of taking photos, people may not fully expose their faces due to the camera angle, light, mood, and equipment. In order to detect faces in different positions or different angles in the image, it needs to first change the scale of the sliding window or the input image. The sliding window also needs to scan the input image according to a certain step size ( $w$ ); as  $w$  increases, the quantity of judgment windows and the amount of computation in the model network will decrease geometrically. When the image is input into the DCNN, there is no need to preprocess the image. However, in order to better detect faces of different sizes, the subacquisition rate of the original input image is scaled at different scales, and then, the image pyramid is obtained, as shown in Figure 4.

In this paper, DCNN is used to process each face image located in the image pyramid according to the input order, and the potential regions in the detected face image are marked at the same time. The coordinate points and size of the specific area of the face in the image are recorded. After all images are processed, all regions of response points in the detected images are reversely mapped to the original face image input in the input layer. Then, the overlapping areas of the image are fused to obtain the final result. Subsequently, face straightening is performed at the stage of no face

alignment, and the face features in different poses are extracted, mainly partial images around the eyes, corners of the mouth, and the tip of the nose in the face image, as shown in Figure 5.

Subsequently, the Euclidean distance [21] and the learning feature matching algorithm are applied to feature matching in the face image, and the matching process is shown in Figure 6.

In the first step, it is necessary to quickly judge the sample that is easy to judge in matching face image samples. The Euclidean distance is used for the similarity between samples, and then, the absolute distance between two points in space is calculated as follows:

$$d(u, v) = \sqrt{\sum_{i=1}^n (u_i - v_i)^2}. \quad (10)$$

In the second step, on the premise of ensuring the feature recognition rate in the face image, a learning algorithm with a higher recognition rate is used to model the features in the face image, and the similarity of the complete face features is matched to realize the refined judgment of the sample. It is evident from Figure 5 that it is needed to extract five local value features in the face image and obtain a total of six 160-dimensional feature vectors, which are spliced into 960-dimensional facial feature vectors. Then, the feature image is reduced to 160 dimensions again by PCA dimension reduction.

Besides, in the second step, joint Bayesian is used to calculate the similarity between samples. Assuming that the two faces are the joint distribution of  $x_1$  and  $x_2$ , respectively, and both obey the Gaussian distribution feature of 0 mean, then the mathematical expression of the covariance between these two face images is as follows:

$$\text{cov}(u_i, v_j) = \text{cov}(\mu_i, \mu_j) + \text{cov}(\varepsilon_i, \varepsilon_j), \quad (11)$$

in which  $\mu$  is the identity of the character and  $\varepsilon$  is the difference between the face itself (light, expression, and posture).

**3.3. Evaluation Indexes for the Testing and Identification of Deep Reconstruction Network Algorithm.** In this paper, Matlab software is used to simulate the deep reconstruction network, and LFW dataset is used for training and testing. In this paper, the following experiments are carried out on TensorFlow platform with Linux as the operating system and 2 \* TITAN X as the GPU. The LFW dataset has a total of 13233 face images, each of which gives the corresponding name, there are a total of 5749 people, and most people have only one picture. The size of each image is  $250 \times 250$ , most of which are color images, but there are also a few black and white face pictures. In this paper, the face images in the dataset are randomly divided into a training set and a verification set. The training set contains 10,000 face images, and the verification set contains 3233 face images. When training the model, first the effects of different parameter settings (learning rate  $\eta$ , attenuation coefficient  $\lambda$ , training



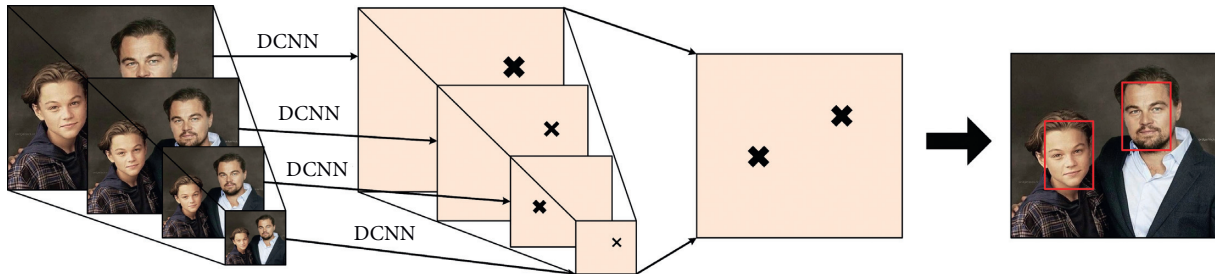


FIGURE 4: Flow chart of face detection in face image.

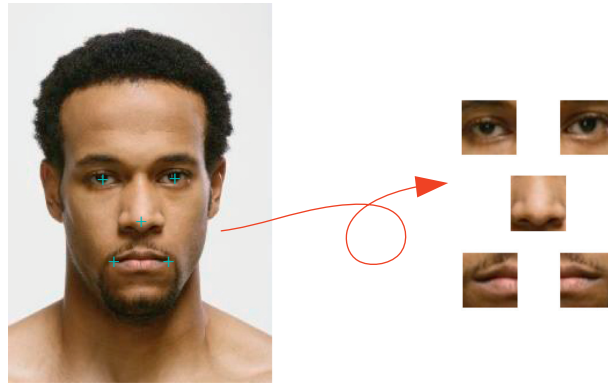


FIGURE 5: Extraction of face feature points and subfigures of each area.

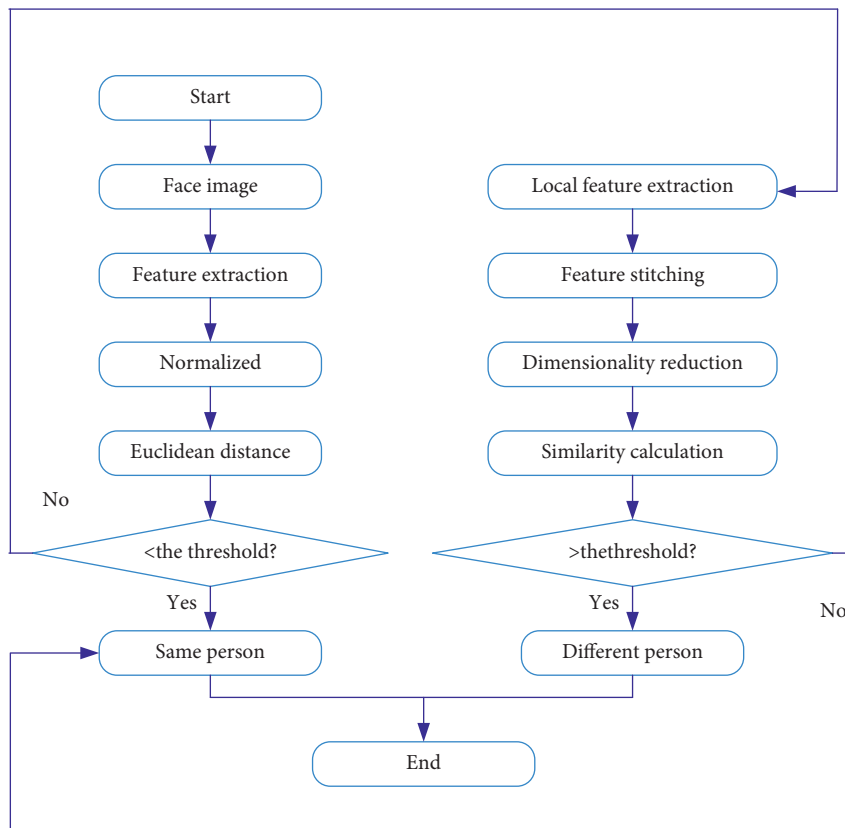


FIGURE 6: Matching process of face feature points.

optimization algorithm, and dropout) on the training effect of the model are analyzed, and the loss value, accuracy value, overfitting rate, and training time are adopted for evaluation of the training effect. Based on the model with the optimal parameter settings, it is applied to the verification set, and the mean average precision (mAP) value is used to evaluate the model recognition effect.

In order to evaluate the effect of the proposed algorithm on face FE, 10140 images of 1150 people from the LFW dataset are selected, and 180 pairs of positive face samples and 180 pairs of negative face samples are randomly selected as test sample pairs. Then, the construction network is used to perform face FE, and the simplest Euclidean distance is used to match feature similarity. Finally, classification accuracy and matching accuracy are used to evaluate the effectiveness of face FE. Subsequently, these 10140 images are used to continue segmentation. Images of 1050 people are selected for facial feature matching algorithm training. For the remaining 100 images of individuals, 200 pairs of positive samples and 200 pairs of negative samples are randomly selected to verify. The recognition accuracy rate, positive sample error rate, negative sample error rate, pairing correct rate, and pairing average time are used to evaluate the matching effect of face features.

In the field of target recognition and detection and recognition, mAP is often used to evaluate the effect of algorithm detection and recognition [22]. In the paper, basic concepts such as error rate, accuracy, precision, recall, and average precision are first introduced. The error rate and accuracy rate are the most commonly used measurement terms in classification evaluation, where the error rate is the ratio of error samples to the total sample, and the accuracy rate is the ratio of correct sample to the total sample. Assuming that the sample dataset is  $S$ , the sample is defined as  $u_i$ , the type of the sample is  $v_i$ , the algorithm's prediction result is  $f(u_i)$ , and then the calculation equations of error rate and accuracy rate are as follows:

$$\text{Error}(f; S) = \frac{1}{m} \sum_{i=1}^m I(f(u_i) \neq v_i), \quad (12)$$

$$\text{Acc}(f; S) = \frac{1}{m} \sum_{i=1}^m I(f(u_i) = v_i) = 1 - \text{Error}(f; S). \quad (13)$$

The calculation equations for the precision rate (Pre) and the recall rate (Recall) are as follows:

$$\text{Pre} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (14)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (15)$$

where TP is the true positive, FP is the false positive, and FN is the false negative.

Using the precision rate as the vertical axis and the recall rate as the horizontal axis, a P-R curve is obtained. The area under the P-R curve is average precision (AP), and mAP is the average of AP indicators in each type of classification

task. In addition, the overfitting ratio (OR) is used to evaluate the algorithm. The calculation equation of OR is as follows:

$$\text{OR} = \frac{\text{TrainAcc}}{\text{ValAcc}}, \quad (16)$$

in which TrainAcc is the training accuracy rate and ValAcc is the verification accuracy rate.

## 4. Results

*4.1. Comparison of Recognition Rate Based on Deep Reconstruction Network Algorithm.* By adjusting the learning rate  $\eta$ , attenuation coefficient  $\lambda$ , training method, and dropout, the impact of different parameters on the accuracy of DCNN model recognition is explored, the database is used to train the DCNN model, and the maximum number of iterations is set to 1000. In the experiment, the activation function in the network is always kept as ReLU. First, the effects of different learning rates  $\eta$  on the model recognition rate are compared. The results are shown in Figures 7 and 8. It is evident from Figure 7 that when  $\eta = 0.0004$ , the recognition accuracy of the algorithm is the highest; when  $\eta = 0.001$ , the recognition accuracy of the algorithm is the lowest. It is evident from Figure 8 that when  $\eta = 0.0004$ , the loss value of the algorithm is the smallest, and when  $\eta = 0.001$ , the loss value of the algorithm is the largest. With the gradual increase in  $\eta$ , the accuracy value gradually decreases and the loss value gradually increases.

Then, the effects of different learning rates on the face recognition performance of the algorithm are quantitatively compared. It is evident from Table 2 that when  $\eta = 0.0004$ , the algorithm has the largest accuracy value (99.03%), the smallest loss value (0.047), and the shortest training time (352 s). When  $\eta = 0.0007$ , the algorithm has the lowest overfitting rate (1.005). In summary, when  $\eta = 0.0004$ , the performance of the algorithm is the best, so follow-up tests are conducted based on this.

The effect of different attenuation coefficients  $\lambda$  on the model recognition rate is compared. The results are shown in Figures 9 and 10. It is evident from Figure 9 that when  $\lambda = 0.0001$ , the recognition accuracy of the algorithm is the highest; when  $\lambda = 0.09$ , the recognition accuracy of the algorithm is the lowest. It is evident from Figure 10 that when  $\lambda = 0.0001$ , the loss value of the algorithm is the smallest; when  $\lambda = 0.09$ , the loss value of the algorithm is the largest. With the gradual increase in  $\lambda$ , the accuracy value gradually decreases and the loss value gradually increases.

Then, the effects of different attenuation coefficients on the face recognition performance of the algorithm are quantitatively compared. It is evident from Table 3 that when  $\lambda = 0.0001$ , the accuracy value of the algorithm is the largest (99.03%), and the loss value is the smallest (0.047). When  $\lambda = 0.0007$ , the training time of the algorithm is the shortest (342 s). When  $\lambda = 0.0001$  and 0.0009, the overfitting rate of the algorithm is the smallest (1.006). In summary, when  $\lambda = 0.0001$ , the performance of the algorithm is the best, so follow-up tests are conducted based on this, and the classifier

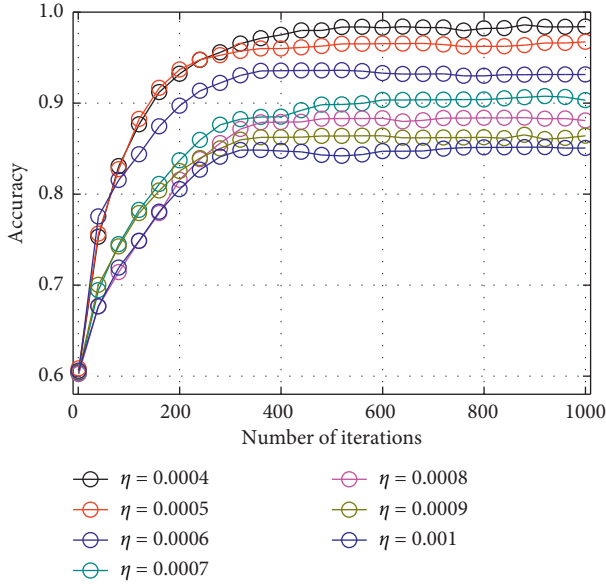


FIGURE 7: Comparison of recognition accuracy of different learning rates.

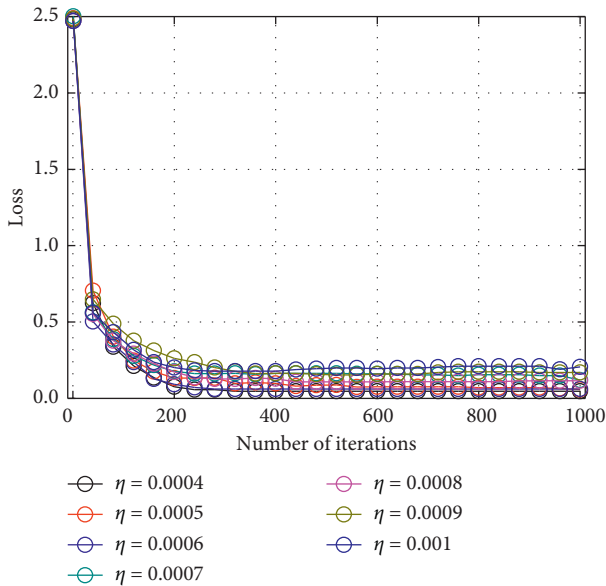


FIGURE 8: Comparison of training loss value of different learning rates.

TABLE 2: Effects of different learning rates on the training effect of deep reconstruction network algorithm.

| $\eta$ | Acc (%) | Loss  | Training time (s) | Overfitting rate |
|--------|---------|-------|-------------------|------------------|
| 0.0004 | 99.03   | 0.047 | 352               | 1.006            |
| 0.0005 | 96.74   | 0.059 | 354               | 1.006            |
| 0.0006 | 93.15   | 0.062 | 368               | 1.082            |
| 0.0007 | 90.31   | 0.120 | 370               | 1.005            |
| 0.0008 | 88.07   | 0.115 | 400               | 1.011            |
| 0.0009 | 86.41   | 0.170 | 398               | 1.026            |
| 0.001  | 85.05   | 0.206 | 405               | 1.058            |

used in this model is determined to be the triplet loss function.

After comparing the effects of different training methods on the model recognition rate, the results are shown in Figures 11 and 12. It is evident from Figure 11 that when the training method is SGD, the recognition accuracy of the algorithm is the highest; when the training method is Adagrad, the recognition accuracy of the algorithm is the lowest. It is evident from Figure 12 that when the training method is SGD, the loss value of the algorithm is the smallest; when the training method is Adagrad, the loss value of the algorithm is the largest.

Then, the effects of different training methods on the face recognition performance of the algorithm are quantitatively compared. It is evident from Table 4 that when the training method is SGD, the algorithm has the largest Acc value (99.03%), the smallest loss value (0.047), and the lowest overfitting rate (1.006). When the training method is RMSprop, the training time of the algorithm is the shortest (351 s). In summary, when the training method is SGD, the performance of the algorithm is the best, so follow-up tests are performed based on this.

Finally, the impact of different dropout values on the model recognition rate is compared. The results are shown in Figures 13 and 14. It is evident from Figure 13 that when dropout = 0.1, the algorithm has the highest recognition accuracy; when dropout = 0.6, the algorithm has the lowest recognition accuracy. It is evident from Figure 14 that when dropout = 0.1, the loss value of the algorithm is the smallest, and when dropout = 0.6, the loss value of the algorithm is the largest. With the gradual increase in the dropout value, the accuracy value of the algorithm gradually decreases, and the loss value gradually increases.

Then, the effects of different dropout values on the face recognition performance of the algorithm are quantitatively compared. It is evident from Table 5 that when dropout = 0.1, the algorithm has the largest Acc value (99.03%), the smallest loss value (0.047), and the lowest overfitting rate with dropout = 0.2 (1.006). When dropout = 0.6, the training time of the algorithm is the shortest (337 s). In summary, when dropout = 0.1, the performance of the algorithm is the best, so follow-up tests are conducted based on this.

It is set that  $\eta = 0.0004$  and  $\lambda = 0.0001$  in the construction of the deep reconstruction network, SGD is selected as the training method, and dropout = 0.1. Then, the effects of different activation functions on the model recognition rate are compared, as shown in Figures 15 and 16. It is evident from Figure 15 that the recognition accuracy rate is the highest when the activation function is ReLu, and the recognition accuracy rate is the lowest when the activation function is sigmoid. From Figure 16, it is evident that the loss value is the lowest when the activation function is ReLu, and the loss value is the highest when the activation function is sigmoid.

Then, the effects of different activation functions on the face recognition performance of the algorithm are quantitatively compared. It is evident from Table 6 that when the

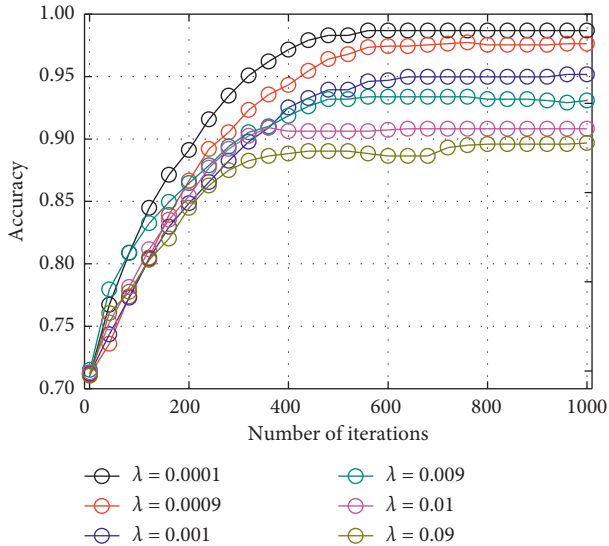


FIGURE 9: Comparison of recognition accuracy of different attenuation coefficients.

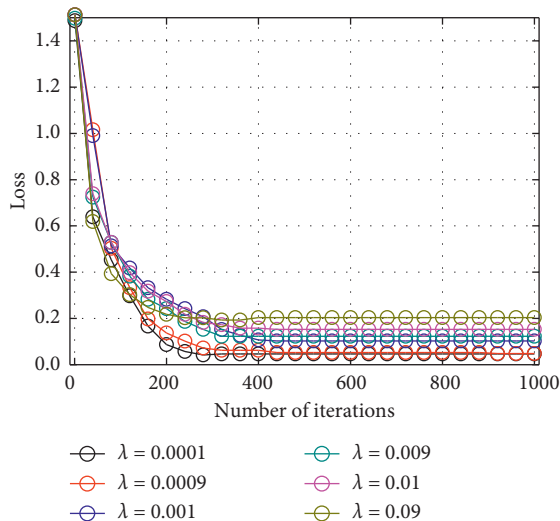


FIGURE 10: Comparison of training loss value of different attenuation coefficients.

TABLE 3: Influence of different attenuation coefficients on the training effect of deep reconstruction network algorithm.

| $\lambda$ | Acc (%) | Loss  | Training time (s) | Overfitting rate |
|-----------|---------|-------|-------------------|------------------|
| 0.0001    | 99.03   | 0.047 | 352               | 1.006            |
| 0.0009    | 97.63   | 0.048 | 342               | 1.006            |
| 0.001     | 95.17   | 0.102 | 358               | 1.037            |

activation function is ReLu, the Acc value of this algorithm is the largest (99.03%), the loss value is the smallest (0.047), and the overfitting rate is the lowest (1.006). When the activation function is Tanh, the training time of this algorithm is the shortest (341 s). In summary, when the

activation function is ReLu, the performance of the algorithm is the best, so follow-up tests are carried out on the basis.

Based on the results obtained above, in the deep reconstruction network constructed in this paper,  $\eta = 0.0004$ ,  $\lambda = 0.0001$ , the training method is set to SGD, and dropout = 0.1 for face recognition verification, and the recognition effect is compared with LetNet-5, GoogleNet, AlexNet, VGG-16, VGG-19, and ResNet network. It is evident from Figure 17 that the area under the P-R curve of the algorithm constructed in this paper is the largest; that is, the mAP value is the largest. The mAP values of different models are ranked, and the result is the algorithm proposed in the paper > ResNet > AlexNet > VGG-19 > GoogleNet > VGG-16 > LetNet-5.

4.2. Face Fe and Matching Verification Based on Deep Reconstruction Network Algorithm. The face recognition and FE algorithm proposed in this paper is applied to the LFW dataset to extract and match face features. The results are shown in Figure 18. It is evident that the algorithm proposed in this paper can effectively extract the features in the face image and can complete the matching of the same features and then realize the recognition and detection of the same face.

Then, the deep reconstruction network algorithm constructed in this paper is applied to FE and feature matching in face images, and the performance is compared with other models. It is evident from Table 7 that when the construction method of this paper is trained, the classification accuracy rate on the training set is 97.94%, and it is higher than the LetNet-5, VGG-16, and VGG-19 algorithms, the accuracy rates of which are 2.1%, 0.94%, and 1.16%, respectively, but lower than GoogleNet, AlexNet, ResNet algorithm, the accuracy rates of which are 0.14%, 0.58%, and 0.46%, respectively. In the verification set, the correct matching rate of the algorithm constructed in this paper is 84.72%, and it is higher than the LetNet-5, VGG-16, and VGG-19 algorithms, the correct matching rates of which are 6.94%, 2.5%, and 1.11%, but lower than the GoogleNet, AlexNet, and ResNet algorithms, the correct matching rates of which are 0.56%, 1.39%, and 1.95%, respectively.

The matching effect of the features extracted from the face image is compared. The comparison results of the total recognition rate are shown in Table 8. It is evident that, after the joint Bayesian method is used in the algorithm proposed in this paper, the correct recognition log rate in the verification set is the highest (88.75%), and it is higher than LetNet-5, GoogleNet, AlexNet, VGG-16, VGG-19, and ResNet algorithms, the correct recognition rates of which are 4%, 1.25%, 0.75%, 3.75%, 3.25%, and 0.25%, respectively.

The effects of using only the joint Bayesian method and the algorithm in this paper on the matching effect of face features are compared, and the results are shown in

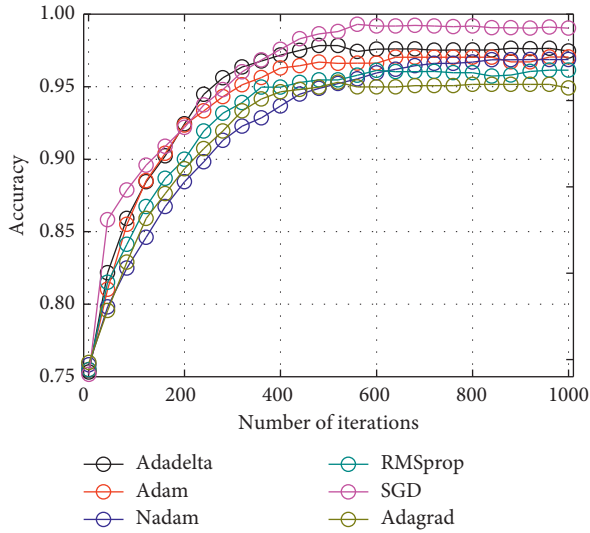


FIGURE 11: Comparison of recognition accuracy of six training methods.

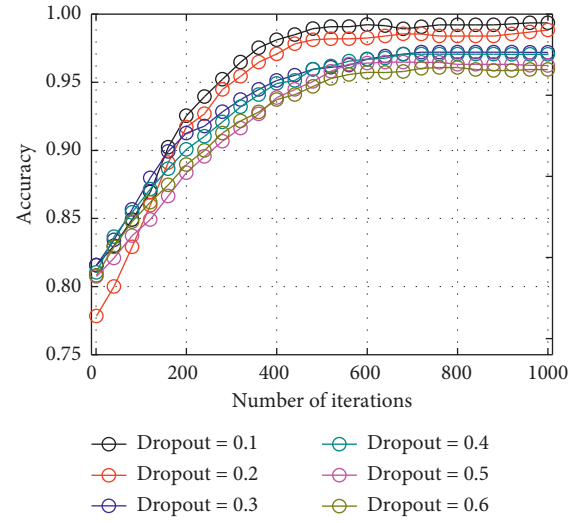


FIGURE 13: Comparison of recognition accuracy with different dropout values.

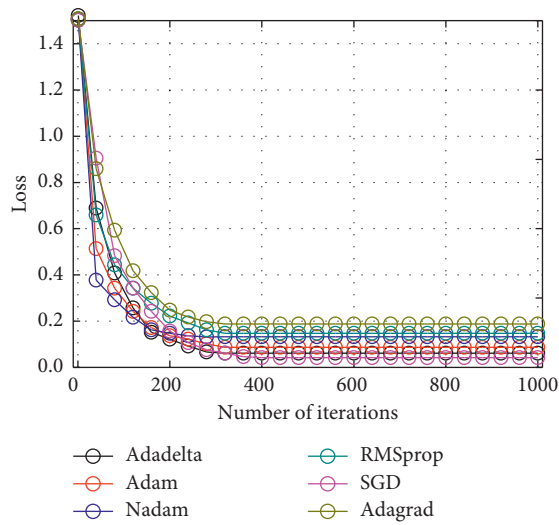


FIGURE 12: Comparison of training loss value of six training methods.

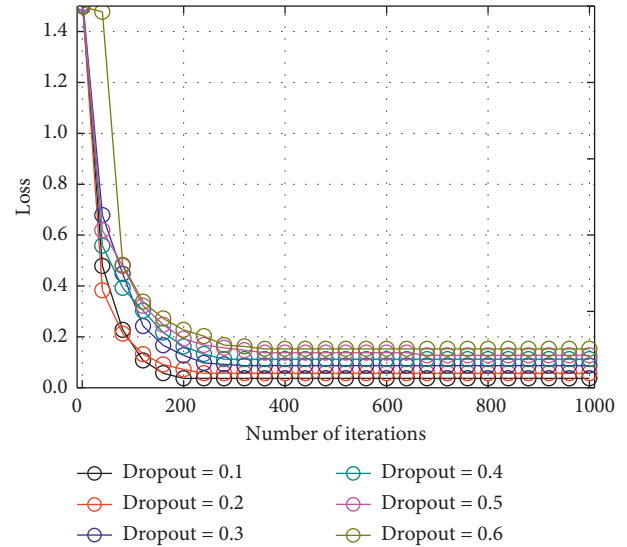


FIGURE 14: Comparison of training loss value with different dropout values.

TABLE 4: Influence of different training methods on the training effect of deep reconstruction network algorithm.

| Training method | Acc (%) | Loss  | Training time (s) | Overfitting rate |
|-----------------|---------|-------|-------------------|------------------|
| Adadelta        | 97.45   | 0.062 | 357               | 1.010            |
| Adam            | 97.05   | 0.087 | 388               | 1.012            |
| Nadam           | 96.87   | 0.132 | 438               | 1.016            |
| RMSprop         | 96.14   | 0.147 | 351               | 1.129            |
| SGD             | 99.03   | 0.047 | 352               | 1.006            |
| Adagrad         | 94.91   | 0.187 | 422               | 1.158            |

Table 9. It is evident that the algorithm proposed in this paper has a faster matching time (206.44 s) and a higher correct matching rate (88.75%) than the joint Bayesian method.

TABLE 5: Effect of different dropout values on the training effect of deep reconstruction network algorithm.

| Dropout | Acc (%) | Loss  | Training time (s) | Overfitting rate |
|---------|---------|-------|-------------------|------------------|
| 0.1     | 99.03   | 0.047 | 352               | 1.006            |
| 0.2     | 98.83   | 0.057 | 347               | 1.006            |
| 0.3     | 97.20   | 0.087 | 348               | 1.009            |
| 0.4     | 97.05   | 0.112 | 351               | 1.017            |
| 0.5     | 96.23   | 0.127 | 340               | 1.027            |
| 0.6     | 95.93   | 0.153 | 337               | 1.518            |

## 5. Discussion

Previous studies have shown that different parameter settings will have a certain impact on the recognition rate of the

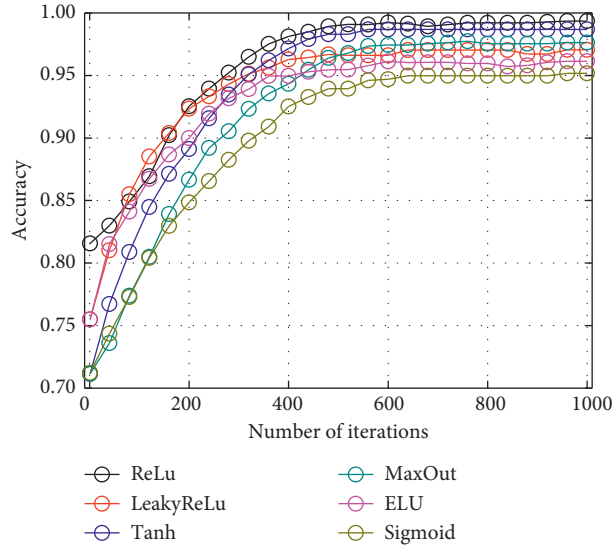


FIGURE 15: Comparison of accuracy of behavior recognition of six activation functions.

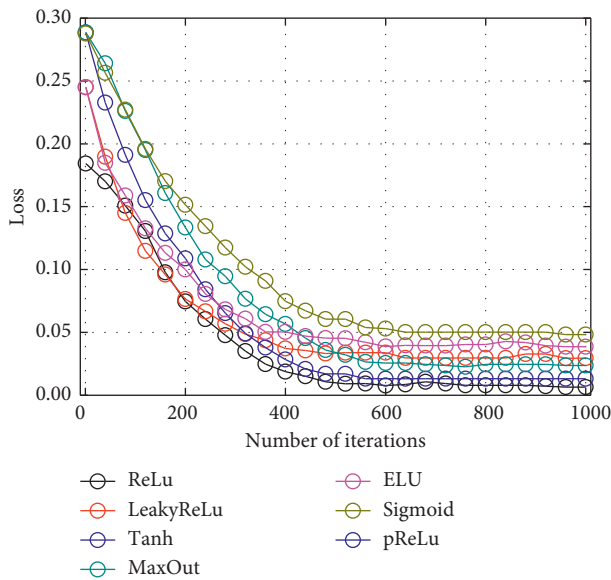


FIGURE 16: Comparison of behavior training loss of six activation functions.

TABLE 6: Influence of different activation functions on the training effect of deep reconstruction network algorithm.

| Activation function | Acc (%) | Loss  | Training time (s) | Overfitting rate |
|---------------------|---------|-------|-------------------|------------------|
| Sigmoid             | 95.17   | 0.091 | 382               | 1.007            |
| ReLu                | 99.03   | 0.047 | 352               | 1.006            |
| Tanh                | 98.52   | 0.053 | 341               | 1.007            |
| LeakyReLu           | 97.11   | 0.073 | 350               | 1.006            |
| ELU                 | 96.78   | 0.087 | 349               | 1.007            |
| MaxOut              | 97.45   | 0.082 | 367               | 1.006            |

DCNN model. In this paper, the effects of different learning rates, training methods, attenuation coefficients, and dropout on the recognition performance of the algorithm are compared. Studies have shown that too large learning rate will cause the algorithm to appear unstable during the learning process, which in turn affects the performance of the algorithm [23], which is consistent with the results of this paper that the learning rate is negatively correlated with the performance of the algorithm proposed in this paper. Then, the effects of different training methods on the recognition rate of the DCNN model are compared. The RMSprop-based network training method will increase the probability of overfitting, and the madam-based network training method will increase the momentum parameter of Nesterov, so the training will take more time. Previous studies have shown that adjusting dropout discarding rate in many network parameters can solve the problem of overfitting in the model, but excessively increasing dropout discarding rate will reduce the RA of image classification, and the overfitting probability will increase [24], which is basically consistent with the results in this paper that the dropout is negatively correlated with the recognition performance of the constructed face recognition algorithm. At the same time, the effect of different attenuation coefficient settings on the recognition performance of the proposed algorithm is compared. It is found that the attenuation coefficient is negatively correlated with the recognition rate of the proposed algorithm. This is because excessive weight attenuation coefficient will destroy the stability of the algorithm in the learning process, so choosing the smallest possible attenuation coefficient can ensure the stability of the face recognition algorithm [25].

Facial features are mainly divided into color (skin color), contour (ellipse), illumination (hair, eyes, and jaw), template

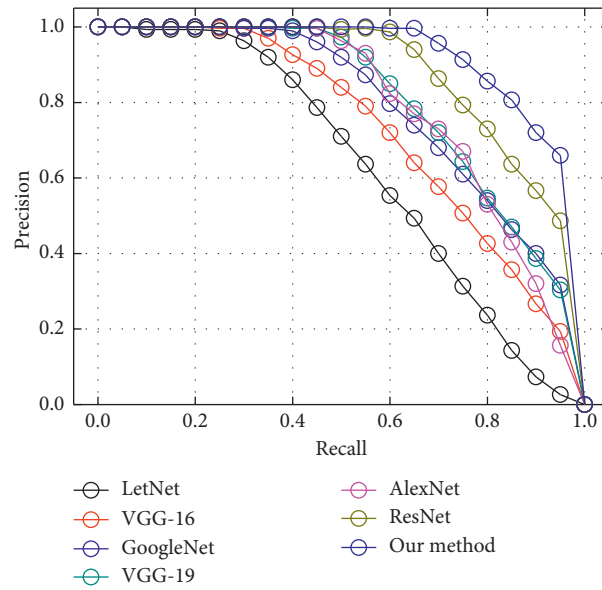


FIGURE 17: Comparison of P-R curves of face recognition performance of different models.



FIGURE 18: Face image FE and matching. Different colored circles in the same picture indicate different extracted features, and the same color in different pictures in the same group indicates matching similar features.

TABLE 7: Comparison of facial FE performance of different models.

| Algorithm  | Correct number of training classifications | Training classification accuracy (%) | Verify correct log | Verification correct match rate (%) |
|------------|--|--------------------------------------|--------------------|-------------------------------------|
| LetNet-5   | 9028                                       | 95.84                                | 280                | 77.78                               |
| VGG-16     | 9117                                       | 96.78                                | 296                | 82.22                               |
| VGG-19     | 9137                                       | 97.00                                | 301                | 83.61                               |
| GoogleNet  | 9239                                       | 98.08                                | 307                | 85.28                               |
| AlexNet    | 9281                                       | 98.52                                | 310                | 86.11                               |
| ResNet     | 9269                                       | 98.40                                | 312                | 86.67                               |
| Our method | 9226                                       | 97.94                                | 305                | 84.72                               |
| LetNet-5   | 9028                                       | 95.84                                | 280                | 77.78                               |

TABLE 8: Comparison of total recognition rates of facial features of different models.

| Algorithm  | Identify the correct logarithm | Recognition correct log rate (%) | Log error rate of positive samples (%) | Logarithmic error rate of negative samples (%) |
|------------|--------------------------------|----------------------------------|--|--|
| LetNet-5   | 339                            | 84.75                            | 13.5                                   | 17   |
| VGG-16     | 340                            | 85                               | 14.5                                   | 15.5   |
| VGG-19     | 342                            | 85.5                             | 15                                     | 14   |
| GoogleNet  | 350                            | 87.5                             | 14                                     | 11   |
| AlexNet    | 352                            | 88                               | 11.5                                   | 12.5   |
| ResNet     | 354                            | 88.5                             | 10.5                                   | 12.5   |
| Our method | 355                            | 88.75                            | 11                                     | 11.5   |

TABLE 9: Comparison of total recognition rates of facial features of different models.

| Algorithm    | Matching time (s) | Match log correctly | Correct match rate (%) |
|--------------|-------------------|---------------------|------------------------|
| United Bayes | 240.15            | 350                 | 87.5                   |
| Our method   | 206.44            | 355                 | 88.75                  |

(mean and variance), transformation domain (feature representation), structure (facial symmetry), inlay (mosaic rule), and histogram features (grey distribution) [26]. In the process of face feature detection, because of factors such as highly nonlinear distribution (such as facial expression and color difference), ornaments (beard, glasses, and hats), the expression (facial muscle movement), light (brightness and angles), image quality (resolution), and a complex scenario (face number and gap), classification errors are caused [27]. Xu et al. proposed a semisupervised method for FR in LFW and YTF databases and found that the recognition rate of the algorithm was 98.63% and 91.76%, respectively [28]. In this paper, the constructed deep reconstruction network is applied to the FE of face images in LFW data; after the proposed algorithm is compared to LetNet-5 [29], VGG-16 [30], and VGG-19 [31] algorithms, it is found that it has a higher matching accuracy rate but slightly lower than GoogleNet [32], AlexNet [33], and ResNet [34] algorithms. However, the algorithm proposed in this paper can reduce the complexity of model generation, while preventing the occurrence of overfitting. Finally, the Bayesian method [35] is used to match face features. It is found that the accuracy rate of the proposed algorithm for face features matching is higher than LetNet-5, VGG-16, VGG-19, GoogleNet, AlexNet, and ResNet. At the same time, compared with the joint Bayes method alone, it can reduce the time consumption of feature

matching [36] and improve the matching accuracy [37], indicating that the algorithm proposed in this paper can quickly perform large-scale face image feature matching and has certain advantages compared with other DCNN models.

## 6. Conclusion

In this paper, a new deep reconstruction network is constructed using the Inception structure in the GoogleNet network and the residual structure in the ResNet network, and it is applied to face recognition for optimal parameter selection of the algorithm. Based on this algorithm, face pyramid and local feature segmentation are applied to construct a face FE algorithm. It is found that it has a better FE effect than the general DCNN model. Finally, based on the algorithm, the face feature matching is achieved using the joint Bayesian method, and the results are verified. The results show that the model constructed in this paper can effectively identify face images with different interference factors. However, the model is trained and verified only through the images in the database. A specific FR system needs to be developed to explore its application effect in video portrait recognition. In conclusion, the results help for the subsequent development of FR system based on DCNN and the improvement of the efficiency of FR.



## Data Availability

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also form part of an ongoing study.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] E. Viccaro, E. Sands, and C. Springer, "Spaced retrieval using static and dynamic images to improve face-name recognition: alzheimer's dementia and vascular dementia," *American Journal of Speech-Language Pathology*, vol. 28, no. 3, pp. 1184–1197, 2019.
- [2] L. Li, P. L. Correia, and A. Hadid, "Face recognition under spoofing attacks: countermeasures and research directions," *IET Biometrics*, vol. 7, no. 1, pp. 3–14, 2018.
- [3] R. Starrfelt, S. K. Klargaard, A. Petersen, and C. Gerlach, "Reading in developmental prosopagnosia: evidence for a dissociation between word and face recognition," *Neuropsychology*, vol. 32, no. 2, pp. 138–147, 2018.
- [4] H. Qin and M. A. El-Yacoubi, "Deep representation-based feature extraction and recovering for finger-vein verification," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 8, pp. 1816–1829, 2017.
- [5] Z. Wang, C. Song, and T. Chen, "Deep learning based monitoring of furnace combustion state and measurement of heat release rate," *Energy*, vol. 131, pp. 106–112, 2017.
- [6] O. Russakovsky, J. Deng, H. Su et al., "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.
- [7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 23, no. 2, pp. 1097–1105, 2012.
- [8] K. Seeliger, M. Fritsche, U. Güçlü et al., "Convolutional neural network-based encoding and decoding of visual object recognition in space and time," *NeuroImage*, vol. 180, pp. 253–266, 2017.
- [9] D. T. Hoang and H. J. Kang, "Rolling element bearing fault diagnosis using convolutional neural network and vibration image," *Cognitive Systems Research*, vol. 53, pp. 42–50, 2018.
- [10] M. Zhang, S. Khan, and H. Yan, "Deep eigen-filters for face recognition: feature representation via unsupervised multi-structure filter learning," *Pattern Recognition*, vol. 100, p. 107176, 2020.
- [11] G. Goswami, A. Agarwal, N. Ratha, R. Singh, and M. Vatsa, "Detecting and mitigating adversarial perturbations for robust face recognition," *International Journal of Computer Vision*, vol. 127, no. 6–7, pp. 719–742, 2019.
- [12] K. Isogawa, T. Ida, T. Shiodera et al., "Deep shrinkage convolutional neural network for adaptive noise reduction," *IEEE Signal Processing Letters*, vol. 25, no. 2, pp. 224–228, 2017.
- [13] T. Nakazawa and D. V. Kulkarni, "Anomaly detection and segmentation for wafer defect patterns using deep convolutional encoder-decoder neural network architectures in semiconductor manufacturing," *IEEE Transactions on Semiconductor Manufacturing*, vol. 32, no. 2, pp. 250–256, 2019.
- [14] F.-P. An and X.-W. Zhou, "BEMD-SIFT feature extraction algorithm for image processing application," *Multimedia Tools and Applications*, vol. 76, no. 11, pp. 13153–13172, 2017.
- [15] S. P. Xu, T. Liu, G. Z. Zhang et al., "A two-stage noise level estimation using automatic feature extraction and mapping model," *IEEE Signal Processing Letters*, vol. 26, no. 1, pp. 179–183, 2018.
- [16] B. Fielding and L. Zhang, "Evolving image classification architectures with enhanced particle swarm optimisation," *IEEE Access*, vol. 6, pp. 68560–68575, 2018.
- [17] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2016.
- [18] E. Kang, W. Chang, J. Yoo, and J. C. Ye, "Deep convolutional framelet denoising for low-dose ct via wavelet residual network," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1358–1369, 2018.
- [19] D. S. Trigueros, L. Meng, and M. Hartnett, "Enhancing convolutional neural networks for face recognition with occlusion maps and batch triplet loss," *Image and Vision Computing*, vol. 79, 2018.
- [20] J. Alan and F. Guo, "Spectral-spatial feature extraction and classification by ann supervised with center loss in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 3, pp. 1755–1767, 2019.
- [21] X. Gan and Z. Haitao, "A noise-robust semi-supervised dimensionality reduction method for face recognition," *Optik International Journal for Light & Electron Optics*, vol. 157, pp. 858–865, 2018.
- [22] U. D. Dixit and M. S. Shirdhonkar, "Face-based document image retrieval system," *Procedia Computer Science*, vol. 132, pp. 659–668, 2018.
- [23] H. Zhao, F. Liu, H. Zhang, and Z. Liang, "Research on a learning rate with energy index in deep learning," *Neural Networks*, vol. 110, pp. 225–231, 2019.
- [24] G. H. D. Rosa, P. P. João, and X. S. Yang, "Handling dropout probability estimation in convolution neural networks using metaheuristics," *Soft Computing*, vol. 4, pp. 6147–6156, 2017.
- [25] M. Tzelepi and A. Tefas, "Improving the performance of lightweight cnns for binary classification using quadratic mutual information regularization," *Pattern Recognition*, vol. 106, p. 107407, 2020.
- [26] T. Amina, A. T. Ahmed, S. Djamel et al., "Deep learning features for robust facial kinship verification," *IET Image Process*, vol. 12, no. 12, pp. 2336–2345, 2018.
- [27] M. Kraemer, Q. B. Huynh, D. Wiczorek et al., "Distinctive facial features in idiopathic Moyamoya disease in Caucasians: a first systematic analysis," *PeerJ*, vol. 6, no. 4, Article ID e4740, 2018.
- [28] W. Xu, J. Wu, S. Ding, L. Lian, and H. Chao, "Enhancing face recognition from massive weakly labeled data of new domains," *Neural Processing Letters*, vol. 49, no. 3, pp. 939–950, 2019.
- [29] D. Wang, H. Lu, and M.-H. Yang, "Kernel collaborative face recognition," *Pattern Recognition*, vol. 48, no. 10, pp. 3025–3037, 2015.
- [30] F. Zhu, X. Kong, H. Fu, and Q. Tian, "Pseudo-positive regularization for deep person re-identification," *Multimedia Systems*, vol. 24, no. 4, pp. 477–489, 2018.
- [31] M. F. Hansen, M. L. Smith, L. N. Smith et al., "Towards on-farm pig face recognition using convolutional neural networks," *Computers in Industry*, vol. 98, pp. 145–152, 2018.
- [32] N. Bi, J. Chen, and J. Tan, "The handwritten Chinese character recognition uses convolutional neural networks with the googlenet," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 33, no. 11, p. 1940016, 2019.

- [33] A. Aslam, K. Hayat, A. I. Umar et al., “Wavelet-based convolutional neural networks for gender classification,” *Journal of Electronic Imaging*, vol. 28, no. 1, p. 1, 2019.
- [34] Z. Lu, X. Jiang, and A. Kot, “Deep coupled resnet for low-resolution face recognition,” *IEEE Signal Processing Letters*, vol. 25, no. 4, pp. 526–530, 2018.
- [35] Z. Huang, X. Xu, J. Ni, H. Zhu, and C. Wang, “Multimodal representation learning for recommendation in internet of things,” *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10675–10685, 2019.
- [36] M. Zhang, D. Zhang, H. Yao, and K. Zhang, “A probabilistic model of human error assessment for autonomous cargo ships focusing on human–autonomy collaboration,” *Safety Science*, vol. 130, Article ID 104838, 2020.
- [37] F. Hu and G. Wu, “Distributed error correction of EKF algorithm in multi-sensor fusion localization model,” *IEEE Access*, vol. 8, pp. 93211–93218, 2020.

## Research Article

# E-Commerce Enterprise Supply Chain Cost Control under the Background of Big Data

Haijun Mao<sup>1</sup> and Long Chen<sup>2</sup> 

<sup>1</sup>School of Economics and Management, Dalian Minzu University, Dalian 116650, Liaoning, China

<sup>2</sup>School of Management, Hebei GEO University, Shijiazhuang 050031, Hebei, China

Correspondence should be addressed to Long Chen; [chenlong85@hgu.edu.cn](mailto:chenlong85@hgu.edu.cn)

Received 23 October 2020; Revised 22 December 2020; Accepted 2 January 2021; Published 16 February 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Haijun Mao and Long Chen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Since the twentieth century, it has been an era of rapid development of information technology; the scale of data is almost the growth rate of the blowout type; no matter what it is, a large number of enterprises or departments are increasing a large number of cost data. However, the current cost management model still remains in the traditional management method and lacks a smarter big data analysis method. In addition, there is a lot of research on big data applications, and there are few e-commerce supply chains. Therefore, the research purpose of this study is to use big data technology to explore a series of practical operation methods for supply chain Cultural Communication Enterprises and summarize the operation mode of building SCC control by using big data technology. In terms of research methods, this study combined bibliographic review and empirical analysis, explored cost-based mobile e-commerce (EU) cost control related to big data information, used smart and digital analysis methods to thoroughly analyze CCE business issues from internal and external supply chains, established an e-commerce business supply chain cost control model based on big data technology and elaborated cost control procedures and measures. Finally, it summarized the research results and drew conclusions to provide a theoretical basis for promoting enterprises products products to reduce supply chain costs. The research in this study has achieved a breakthrough in the cost management and control of EE; it provides empirical guidance and theoretical reference for EE to adopt big data technology for cost command of supply chain (CCSC), could help EE to reduce cost of supply chain management to gain higher profit margins, and promote e-commerce industry as a whole to the next level eventually. This study concluded that the use of big data technology for cost command can solve a series of problems effectively, such as the lack of systematic analysis of cost, the lack of contractual partners, the serious waste of sales links, and the policy errors of logistics links, and continuously improve the enterprise management level and the decline of comprehensive cost. The application mode of supply chain CCE enterprises using big data technology constructed in this study has universal applicability.

## 1. Introduction

With the popularity of the computer and the rapid development of Internet, the worldwide data online, platform, and cloud technology, such as vocabulary, more and more frequent in people's daily life, the need for a smartphone has been preliminarily realized. Chinese Internet users have also been broke through 800 million; this is really beyond the forecast process of some experts, can say, today's world has entered the Internet era of big data in advance. The development of big data technology in daily life is not only a subversive change to the traditional way of life but also a

profound change to the production and operation mode of traditional enterprises. Under the background of big data technology era, most enterprises relying on traditional sales means have been seriously impacted, and some enterprises are difficult to adapt to the changes of big data era and survive [1].

More and more people realize that the development of enterprises must take advantage of the trend and only rely on the advancement of current Internet technology to innovate and change the business model and get out of a development model that not only has the traditional offline model but turns to the integration of the traditional offline model and

modern online model [2]. More and more business operators are understood; the economy and the great development of science and technology have brought the big changes in the business mode and not only will have a profound impact on supply and demand; in the context of the overall supply, companies increasingly need to spend more cost to satisfy the consumers with a variety of consumer demand; in order to fundamentally profit space, enterprises urgently need to be done in each link cost control, and the informationization of modern society relying on the traditional cost accounting method efficiency is extremely low and obviously cannot meet the needs of the era of the current business development at high velocity. If enterprises want to change this situation, they must innovate the cost control mode. The arrival of the era of big data makes all this possible. The development of big data technology makes every data full of effective value space, and every piece of information contains a large number of value clues, which undoubtedly provides some enlightenment for many enterprises in transition [3]. How to make full use of Internet big data technology to analyze and control the cost of production and sales and realize enterprise profit has become a major issue that all enterprises need to face and solve.

The current technology for large data in the electronic commerce research mainly concentrated in the following three points: first, in terms of the application of the big data technology, with the influence of the technology, the government pays more and more attention to the technology as it plays a role in economic development; the domestic experts and scholars have put forward their own mentality, and the representative professor Wang Shan from big data technology professional characteristics discusses the data about the role of e-commerce platform to create. He discussed different aspects that infect big data technology and analyzed the technology of data in the foreseeable future opportunities and challenges facing; he thinks that big data technology in the application of e-commerce is a double-edged sword and can better serve the economic construction; it is very difficult, if there is no scientific system to regulate guarantee has been playing a positive role [4]. Second, in terms of the cost command, the scholar Han Lan believes that the cost command of the enterprise should be extended to all related enterprises in the whole supply chain rather than confined to the enterprise itself. He shall, in accordance with the scientific supply chain management thinking, in order to control cost as the core goal to set up the control model of SCC and its cost to enterprises, use the model analyzed with the laws of the cost management thinking and understanding and successfully applying cooperation and game thinking into SCC control thinking [5]. Third, in the cost control and combining uses of big data technology, a relevant expert Cheng Ping with large depth of the data mining technology construct the ERP management model; the model mainly integrates the cloud computing platform, the Internet of things technology, large inventory management, and data analysis center four core modules; e-commerce products were introduced in detail from the inbound to the outbound way of concrete application and pointed out the enterprise plays a leading role in this process [6]. By consulting a large amount of literature, we found that

domestic and foreign scholars have basically the same pace of research on this issue, and both of them focus their research on the application field, cost control, and the combination of the two. First, in terms of the application of large data, Martin Christopher's study, this study discusses the big data technology development potential and future development direction and is trying to expand its application field to healthcare, retail, education, and other industries; he also probes into the technology of big data in the department of public safety space, and the big data technology popularization after how to ensure the security of personal information anxiety is proposed. Second, in terms of SCC command, Charles Bolier proposed SCC management strategies based on the challenges and opportunities of big data [7]. He also focused on analyzing how enterprises should improve the level of cost command by applying the principles of supply chain in the process of cost command [8]. Finally, in the electronic commerce enterprise cost control and combining use of big data technology, Michel Roberto thinks that the future competition between electric business platforms will focus on the logistics speed contest; he thinks target cost law should get electricity business management platform and thinks rising competitiveness comes from the reasonable positioning of the enterprise; the enterprise's distribution network can improve the distribution of costs and benefits, in terms of specific measures, and he also gives some suggestions and opinions to the construction [9, 10].

It is not difficult to find that although the current research on the control of traditional SCCs has become mature, the research focus of many experts, and scholars have not focused on the background of big data technology, and the combination of supply chain ideas and the target cost method has basically achieved an organic combination. However, at present, experts and scholars have not been able to really shift their attention to the background of big data technology. Their current research is still in the traditional medical care, education, and search industries, and there is still a lack of study on the integration in e-commerce, a sunrise industry. Therefore, after fully consulting and analyzing the shortcomings of domestic and foreign research results, this study decides to carry out the following innovations on the existing research: first, the research purpose of this study is clear, which directly targets the innovation of the cost control mode under the background of big data technology. Second, this study boldly broke through the current research is only limited to the theoretical dilemma, the first to carry out specific application research. Finally, this study is neutral and does not have the color of commercial dependence, which is different from previous studies that are often subject to the manipulation of a specific company or organization and affect the objective neutrality of research results.

## 2. Related Research Review and Theoretical Basis

*2.1. Research Practice.* This study is split into five parts, and the command of each part is arranged.

The first part is the referral. In this part, the significance of this research is stated [11]. On the basis of analyzing numerous research results, we summarized the current research status and objectively pointed out the limitations of

existing research. Finally, according to the goal of this research and the idea of improving the existing research, the new innovation point and goal of this research are summarized [12]. The second part of this study is a review of relevant research and a summary of the theoretical basis; mainly in this study, core concepts and basic theories were described, and the SCC command method of this study is described [13, 14]. The third part analyzes the trouble existing in the CCE enterprises. This study analyzes the trouble existing in the cost command process of e-commerce industry. The fourth part is the construction of the cost command model of big data technology in the supply chain of EE. This study adopts the method of case analysis and constructs a scientific cost control application model [15]. The fifth part is the summary of the research conclusion.

## 2.2. Core Concept Elaboration

**2.2.1. Big Data Technology.** Big data is a collection of data processing technologies with a large scale and the ability to store, analyze, and process data for surpassing the scope of conventional database processing capabilities. Big data has fast data flow cycle, massive data scale, low value density space, and diversified data types [16]. Different from traditional database processing technology, big data need to go through the deep processing of big data processing technology to truly play its value. Massive parallel processing, distributed file, data mining, and distributed database technology is the big data technology to adapt to the form, and the concrete form of big data is often characterized by big data pretreatment technology, data acquisition technology, and data statistical analysis technology [17, 18]. Large data pretreatment technology is the combination of large data acquisition technology of large amounts of data input to a centralized distribution of a large database, and on the basis of the data preprocessing and data cleaning, make vast amounts of information data and valuable data, and the characteristic of the process need a lot of import data; its transport amount could reach MB per second or even in the near future to gigabit per second. Data collection technology is to use multiple database tags to collect and store data information from the Internet, mobile phone client, and other end users [19]. The depth of the data mining technology is according to the relevant provisions of the state and the request; using the algorithm classifies the concept of the data and extracting; the technology is mainly suitable for column type storage calculation and may use GreenPlum algorithm; the process of data mining is the most critical step because the link is often needed to use tools and Kmeans clustering Naive Bayes classification and SVM statistical tools to assist in the completion; the challenge to give priority to with single thread control algorithm is complex and both computation and large amount of data [20].

**2.2.2. SCC and Control Theory.** SCC refers to the sum of all costs incurred by enterprises in production, sales, procurement, and other links to support the operation of the supply chain. The cost of supply chain management is the extension of the cost of the whole supply chain. Cost of SCC

control is the enterprise according to the previous data as well as the recent strategic arrangements and make cost management control plan in advance before the creation of the cost or being in the process of producing cost of the various factors that affect the cost to ensure the expected cost goals to achieve management behavior.

The theory of SCC command mainly includes transaction cost theory, interorganization cost management theory, and value chain theory. Transaction cost is the cost generated in the transaction process, including negotiation, contrast, contract signing, supervision performance, incentive, and other expenses. In addition, interest sharing is the basis for long-term cooperation between supply chain companies. It is based on the contribution of each company to the entire supply chain, but it is actually difficult to divide and evaluate the tribute of each enterprise. In this case, the profit distribution may be unfair, resulting in a large number of agency arbitration events, which leads to the increase of transaction costs, and the increase of these costs has no economic benefits. Therefore, only by mutual trust and full information sharing can enterprises reduce the total cost of the supply chain. Interorganizational cost management is to regard the upstream and downstream enterprises or related enterprises as the same organization. The common goal of the organization is to reduce the total cost of the supply chain through joint efforts. All the relevant enterprises in the supply chain share risks and share profits and strive to improve their profits and efficiency on the same ship. The greater the contribution of the chain, the more profit is allocated. However, the interorganizational cost management mode is largely dependent on the coordination among the interorganizational enterprises, so the model is more suitable for information sharing and fine-grained supply chain with larger interaction. The greater the contribution of the chain, the more profit is allocated.

### 2.2.3. Methods of SCC Control

- (1) Target cost control method. The target cost control method is based on customer demand and market-oriented. It uses value engineering to analyze the cost function in the planning and design stages of products and thus creates a method to manage the cost.
- (2) Activity cost control method. Activity-based costing is usually based on the guiding ideology of "products consume activities, activities consume resources." Under the guidance of this idea, resource cost is allocated to the process of operation, and then, the cost of product consumption is tracked based on the driver of activity, so as to calculate the unit cost of various products and the cost of bulk total products to achieve the optimal cost management scheme.
- (3) Responsibility cost control method. Responsibility cost refers to the cost generated in the process of assigning the enterprise to each responsibility management center. The establishment of responsibility cost needs the support of the responsibility

cost system, which establishes the responsibility cost center according to the production and operation structure of the enterprise, divides responsibility into the control cost control effect, and carries out performance assessment, and carries out cost control responsibility for each department and specific executive to control cost.

- (4) Standard cost control methods. The standard cost control method is a comprehensive management method which combines the accounting function, feedback function, and forecast function organically.

The management mode constructed in this study is to distinguish the process steps, including the comprehensive cost management method of pregeneration prediction and adjustment in the process of production and postproduction analysis. In the forecasting stage before cost generation, it is necessary to use target cost law and standard cost law to reasonably forecast the cost of each department. In the process of cost generation, it is often necessary to conduct repeated accounting combined with activity-based cost rule. After the cost is generated, the utility of the cost control process is judged, and the consequence is held accountable according to the liability cost law.

**2.3. Supply Chain Cost Control Methods.** According to the "Organizational Industry Theory," first assume that consumers have the following preferences:

$$U = \int_0^{Q_s} Q_s - P. \quad (1)$$

The positive real number  $Q_s$  is the quality preference parameter of consumers. At a given price, all consumers like high quality, but consumers with high quality preferences are more willing to spend money for high quality. The model assumes that product preferences are uniformly distributed in a positive interval  $(Q_L, Q_H)$ :

$$f(Q) = \frac{1}{\Delta Q} (Q_L < Q < Q_H). \quad (2)$$

Assuming that there is only one product in the market with quality  $s$  and price  $p$ , the demand for this product is equal to the number of consumers who have product preference parameter  $Q$  and satisfy  $Q_s \geq p$ . In other words, assuming that the total number of consumers is 1, the demand for the product is

$$D(p, s) = \begin{cases} 0, & \left(\frac{p}{s} > Q_H\right), \\ \frac{Q_H}{\Delta Q} - \frac{p}{\Delta Q \cdot s}, & \left(Q_L \leq \frac{p}{s} \leq Q_H\right), \\ 1, & \left(\frac{p}{s} < Q_L\right). \end{cases} \quad (3)$$

The demand for products cannot reach the total number of consumers, so it can be seen that the actual demand for products is

$$D(p, s) = \frac{Q_H}{\Delta Q} - \frac{p}{\Delta Q \cdot s}. \quad (4)$$

**2.4. Financial Benefits under Supply Chain Management.** Assume that the supply chain is a simple supply chain consisting of a manufacturer and a seller. Assuming that the quality of the product is constant, the manufacturer and seller only determine the output  $QM$ , the sales  $QR$ , and the final market price  $P$  of the product. The median price  $PM$  of the product is set as the product price that the manufacturer asks for when the product is delivered to the seller's location, which actually includes the production cost and transportation cost of the product. Assume that the manufacturer's marginal production cost and transportation cost of the product are uniformly recorded as  $CM$ , and the seller's marginal cost of sales of the product is  $CR$ . It is assumed that the information between the manufacturer and the seller is symmetrical and complete, that is, the demand function faced by the product and  $CM$  and  $CR$  are the common knowledge of the manufacturer and the seller.

When manufacturers and sellers cooperate to determine output and final prices, the supply chain can be regarded as a "super organization." The purpose of manufacturers and sellers is to maximize the profits of the entire supply chain. Distribute effectively and reasonably among enterprises. Then, the financial benefits of the organization can be seen as

$$\max \pi = (p - CM - CR) * QM = (p - CM - CR) * \left(\frac{Q_H}{\Delta Q} - \frac{p}{\Delta Q \cdot s}\right). \quad (5)$$

$P_0$  is the consumer's maximum payment capacity for the product. Assuming that the payment capacity is large enough to not affect the company's decision-making conditions, the solution can be obtained:

$$\begin{aligned} p^* &= \frac{1}{2} (sQ_H + CM + CR), \\ QM^* &= \frac{1}{2\Delta Q} \left(Q - \frac{CM + CR}{s}\right), \\ \pi^* &= \frac{1}{2s\Delta Q} (sQ_H - CM - CR)^2. \end{aligned} \quad (6)$$

It can be seen that  $\pi^*$  is the largest profit achieved under the collaborative management of the supply chain.

**2.5. Financial Benefits under Nonsupply Chain Coordination.** Assuming that the manufacturer has absolute authority and can unilaterally determine the most favorable intermediate price  $PM$ , since the assumption is that the information is symmetric, the manufacturer can accurately predict the seller's optimal price and optimal sales volume. A rational manufacturer should make own output and sales volume consistent, that is,  $QM = QR$ ; then, the manufacturer's profit function is

$$\begin{aligned} \max \pi &= (CM - CR) * \left( \frac{Q_H}{\Delta Q} - \frac{p}{\Delta Q \cdot s} \right), \\ \text{s.t. } \max_p \pi &= (P - PM - PR) * \left( \frac{Q_H}{\Delta Q} - \frac{p}{\Delta Q \cdot s} \right). \end{aligned} \quad (7)$$

Solving this function can be obtained, and the optimal solution for the seller is

$$\begin{aligned} p^* &= \frac{1}{2} (sQ_H + PM + CR), \\ QM^* &= \frac{1}{2\Delta Q} \left( Q_H - \frac{PM + CR}{s} \right), \\ \pi^* &= \frac{1}{16s\Delta Q} (sQ_H - CM - CR)^2. \end{aligned} \quad (8)$$

For the manufacturer, the optimal solution is

$$\begin{aligned} p^* &= \frac{1}{4} (3sQ_H + CM + CR), \\ QM^* &= \frac{1}{4\Delta Q} \left( Q_H - \frac{CM + CR}{s} \right), \\ \pi^* &= \frac{1}{8s\Delta Q} (sQ_H - CM - CR)^2. \end{aligned} \quad (9)$$

The total profit of the supply chain is

$$\pi_{SC1} = \frac{1}{16s\Delta Q} (sQ_H - CM - CR)^2. \quad (10)$$

Assuming that the seller has absolute power and can unilaterally determine the most favorable middle price PM because the hypothetical information is symmetric, the seller can accurately predict the seller's optimal price and optimal sales volume, and a rational seller should make your sales volume consistent with the expected output, that is,  $QM = QR$ ; then, the seller's profit function is

$$\begin{aligned} \max_p \pi &= (P - PM - CR) * \left( \frac{Q_H}{\Delta Q} - \frac{p}{\Delta Q \cdot s} \right), \\ \text{s.t. } \max_p \pi &= (PM - CM) * \left( \frac{Q_H}{\Delta Q} - \frac{p}{\Delta Q \cdot s} \right). \end{aligned} \quad (11)$$

Solving this function can be obtained, and the optimal solution for the seller is

$$\begin{aligned} p^* &= \frac{1}{2} (sQ_H + CM + CR), \\ QM^* &= \frac{1}{2\Delta Q} \left( Q_H - \frac{CM + CR}{s} \right), \\ \pi^* &= \frac{1}{4s\Delta Q} (sQ_H - CM - CR)^2. \end{aligned} \quad (12)$$

As for the manufacturer, his profit depends entirely on the intermediate price and the order quantity. In this case,

the manufacturer is completely passive, and the manufacturer's profit is 0. The profit of the entire supply chain is

$$\pi_{SC2} = \frac{1}{4s\Delta Q} (sQ_H - CM - CR)^2. \quad (13)$$

Compared with the formula profit under collaborative management, there is no loss, but because the producer has no profit, this structure is very unstable.

### 3. Analysis of the Problems and Causes in the Cost Control Process of the Supply Chain of EE

#### 3.1. Problems and Reasons of External Supply Chain

*3.1.1. Lack of System.* Instead, it makes a preliminary calculation of the costs of product sales, management, operation, and production and processing, which are difficult to truly reflect the logistics cost. Therefore, EE have not yet made accurate cost accounting in terms of logistics costs, and it is difficult to control the costs of each link in detail. In addition, it is difficult to evaluate the marketing impact and understand the true intentions of customers to accurately control them. The imperfect information security of the e-commerce platform is also an important reason for adapting to the difficulties of SCC e-commerce. Information security in the information society is also very important for e-commerce operators. A series of operations of consumers purchasing consumer goods often involve personal data security in various aspects. However, the information security situation of EE is not optimistic at present, and consumers' information such as registration, browsing, payment, purchase, and address are still at risk of being leaked. The EE exposed to the existence of leakage of user information scandal will bring incalculable losses to enterprises.

*3.1.2. Lack of Stable Business Partners.* Purchasing departments are often cheaper than prices as a means of supplier selection at this stage; but even with long-term cooperation, they are not formed for long-term cooperation in supply chain channels. In fact, relationships tend to be due to price discounts and improved quality of service for suppliers, but a truly valuable spirit of cooperation must be based on contracts. EE is always looking for a partner because there is no stable partner. Such frequent partner changes actually increase the cost of inventory management.

Problems existing in the external supply chain have an adverse impact on cost control, which is reflected in the lack of systems and stable business partners. Specific research data are shown in Table 1 and Figure 1.

*3.2. Problems and Reasons of Internal Supply Chain.* The cost analysis results of internal supply chain issues affecting the supply chain are shown in Table 2 and Figure 2.

*3.2.1. Trouble and Cause in the Procurement Process.* Now, many EE do not have a clear definition of the rights and responsibilities of the procurement system, and there

TABLE 1: Cost analysis table of external supply chain problems affecting supply chain.

| External supply chain problems affect SCC analysis table |  |  |                                      |                    |
|--|--|--|--------------------------------------|--------------------|
| Negative composition reduces institutional defects       | High cost operation project name             | Current cost value (ten thousand yuan) | Ideal cost range (ten thousand yuan) | Results the rating |
| System lack  | Imperfect audit system                       | 37                                     | 20-30                                | Not ideal          |
|  | Evaluation system standards are inconsistent | 32                                     | 18-25                                | Not ideal          |
|  | Consumer information security                | 46                                     | 35-50                                | Fair               |
| Lack of stable business partners                         | Procurement costs                            | 18                                     | 10-15                                | Not ideal          |
|  | Cost of finding business partners            | 10                                     | 5-8                                  | Not ideal          |

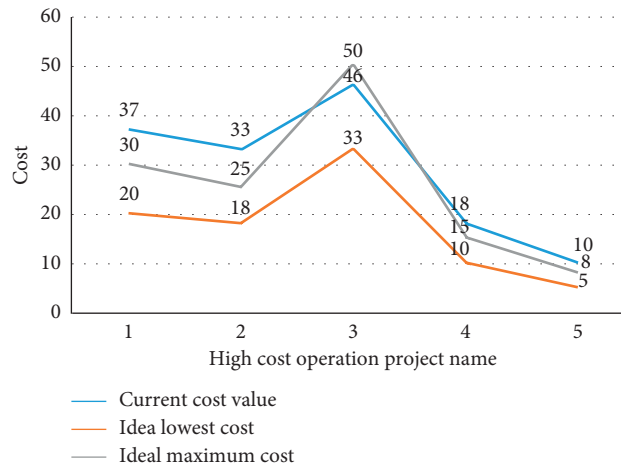


FIGURE 1: Cost analysis table of external supply chain problems affecting supply chain.

are many problems not only in the postsetting, even some nominal, but also in the name of saving costs. For individual purchasers, the lack of audit link or slack audit makes the procurement link prone to fraud. This responsibility setting mode that fails to regulate the procurement standard will inevitably lead to the procurement link cost higher than the actual required cost, resulting in inflated cost.

**3.2.2. Trouble and Cause in the Sales Link.** Now, most large e-commerce platforms lack quantitative segmentation of the sales market, with large span of sales objects and levels, complex business scope, and failure to form core competitiveness, which inevitably leads to excessive marketing costs.

**3.2.3. Existing Problems and Reasons in Logistics.** Along with our country logistics network, the upgrade of the actual logistics cost is gradually narrowing, but the implementation of the phase to the terminal consumers rarely see the substantial fee reduction effect, still appeared in some remote areas, and the present situation of the cost is higher and higher; this seems to increase the income of electricity, but in terms of volume compromised, and network construction is, in fact, is not conducive to reduce logistics cost of recovery.

The side effect of the lack of the internal supply chain system on cost reduction is also increased. The survey and analysis table is given in Table 2.

#### 4. Construction of the Supply Chain Cost Control Instruction Mode under Big Data

Based on the analysis of a huge number of cases, this study constructs a model of using big data technology to control the cost of supply chain. This model is based on the analysis and research of a large number of e-commerce cases through big data technology, which is the core part of this study with most energy.

**4.1. Construction Principles of the SCC Command Model.** Building big data technology under the cost command model is a complex and systematic project, wants to get ideal result, inevitable need in the process of constructing the model following the general principles of control, and in particular also requires a combination of electricity enterprise organization structure, the advantages and disadvantages, the enterprise strategic target, and the industry of the enterprise environmental factors. These factors can be summarized in the following principles: first, to cut the cost as the basic guide; second, pay more attention to the control



TABLE 2: Cost analysis table of internal supply chain problems affecting supply chain.

| External supply chain problems affect SCC analysis table |  |  |                                      |                       |
|--|--|--|--------------------------------------|-----------------------|
| Negative composition reduces institutional defects       | High cost operation project name                                 | Current cost value (ten thousand yuan) | Ideal cost range (ten thousand yuan) | Results the rating    |
| Procurement  | Define the rights and responsibilities of the procurement system | 45                                     | 25–35                                | Not ideal             |
|  | Individual purchaser qualification review                        | 7                                      | 10–12                                | Is not up to standard |
| Sales section  | Quantitative segmentation of sales markets                       | 32                                     | 25–30                                | Not ideal             |
|  | Sales level span   | 21                                     | 15–20                                | Not ideal             |
|  | Business scope span  | 2                                      | 0.5–1                                | Not ideal             |
| Logistics links  | Construction of the logistics network                            | 136                                    | 80–110                               | Excessive             |
|  | Terminal earnings  | 56                                     | 60–100                               | Is not up to standard |

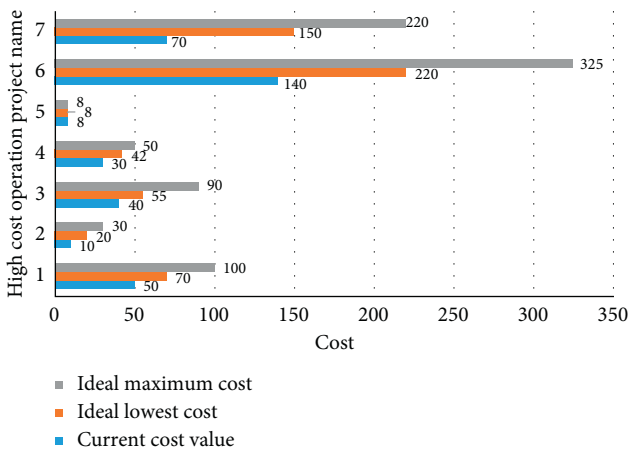


FIGURE 2: Cost analysis table of internal supply chain problems affecting supply chain.

of core costs; third, the implementation of the employee responsibility incentive and punishment system, to ensure the cost of the problem to the people responsible; and fourth, strengthen overall planning for the cost of each link to avoid the domino effect. The priority of these principles is shown in Figure 3.

4.2. *SCC Analysis under Big Data Technology.* SCC under big data technology mainly includes commodity cost, ordering cost, transaction cost, information cost, labor cost, and transaction cost. Among them, the cost incurred in a series of negotiation, contract signing, and negotiation activities with the supplier is the transaction cost. Information cost refers to the cost of unable to effectively select high-quality suppliers caused by information asymmetry. The order cost is divided into fixed cost and variable cost. The order number and variable cost are related, and the fixed cost does not change. The total cost is the sum of information cost, ordering cost, and commodity cost. The cost composition of the procurement link is shown in Table 3 and Figure 4.

### 4.3. Analysis of the Role of the SCC Command Mode

4.3.1. *Continuously Improve Performance Appraisal Results.* The control effect in turn improves performance and helps to continuously improve performance evaluation results. The schematic diagram of the 2016–2020 KPI performance evaluation is shown in Figure 5.

4.3.2. *Continuous Improvement of Enterprise Supply Chain Management.* Prior to cost command, the organization can make a reasonable budget and inform and correct the deviation in time in cost control. After cost control, the course can be summarized, and the next management plan can be improved. Enterprising the cost of data in the business process is improved, through scientific analysis of defects in the process of management, to improve and promote the improvement of the enterprise management level, not only it can produce positive impact to the enterprise, and the progress of the domain also has a strong reference meaning, This field also has strong reference significance. The distribution diagram of the standardized enterprise supply chain management model is shown in Figure 6.

4.3.3. *Steadily Improve the Comprehensive Competitiveness of Enterprises.* EE to strengthen the training and the training of staff, improve staff’s professional skills and culture quality, fundamentally protect the company’s operating efficiency, effectively command the company cost, implement relevant indicators, and make the enterprise overall goals agreed, can promote to achieve the enterprise strategic goals; the standardization of the competitive enterprise image shows the diagram as shown in Figure 7. After the establishment of the cost command system, the cost command can be effectively implemented. With the improvement of the enterprise management level, enterprise competitiveness will be more and more strong. The enterprise will have strong and huge space.

Through the establishment of the assessment system, the innovation of the management level and the improvement of comprehensive competitive power, the cost of EE under the background of big data technology is significantly reduced

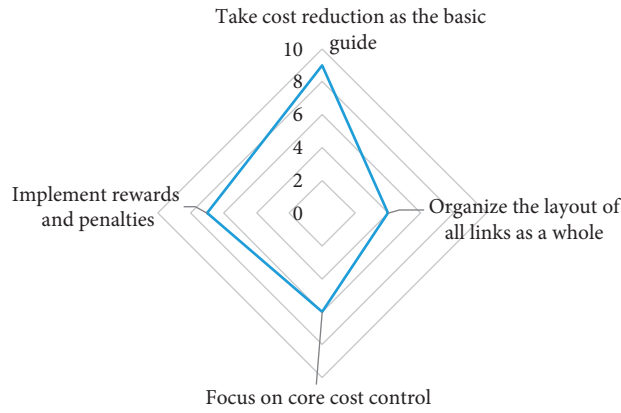


FIGURE 3: Radar chart of influencing factors of SCC control.

TABLE 3: Cost composition chart of the procurement link.

|                               |                       |                      |                       |
|-------------------------------|-----------------------|----------------------|-----------------------|
| Cost structure Percentage (%) | Transaction costs 17% | Their own costs 16%  | Ordering costs 25%    |
| Cost structure Percentage (%) | Artificial costs 17%  | Information costs 8% | Transaction costs 17% |

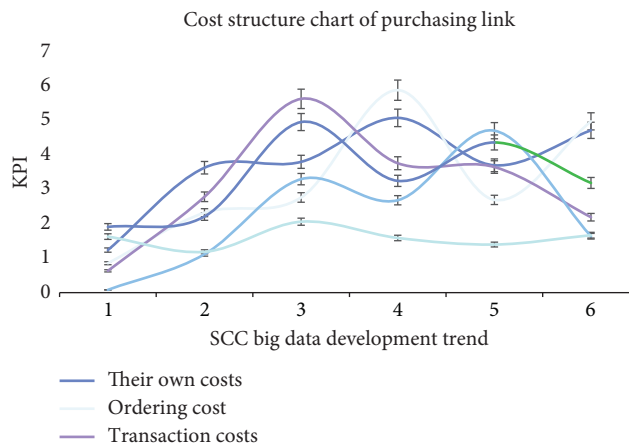


FIGURE 4: Cost composition chart of the procurement link.

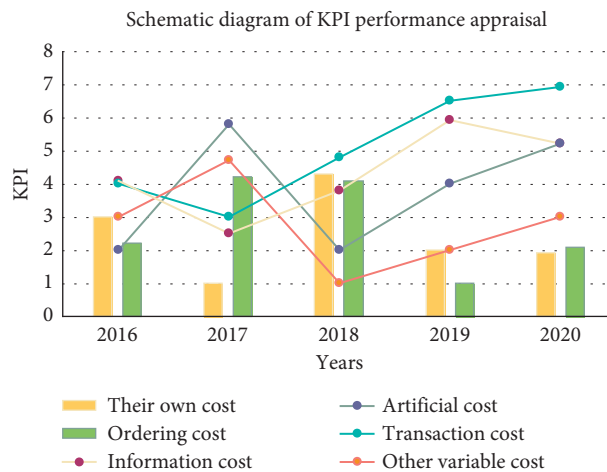


FIGURE 5: Schematic diagram of KPI performance appraisal.

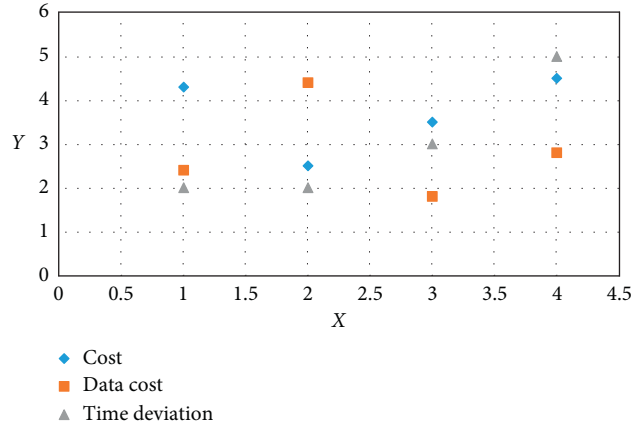


FIGURE 6: Demonstration chart of the standardized management enterprise supply chain level.

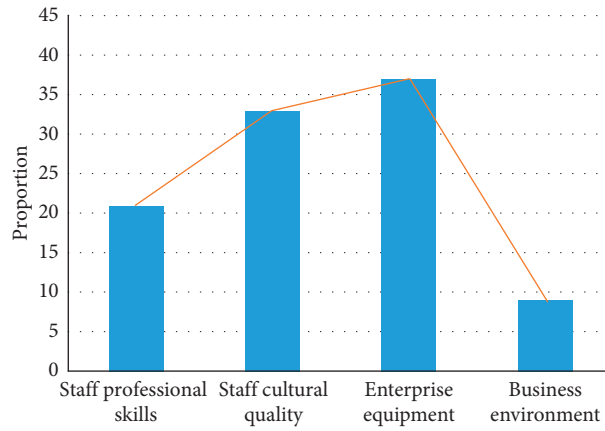


FIGURE 7: The standardized corporate image with strong competitiveness.

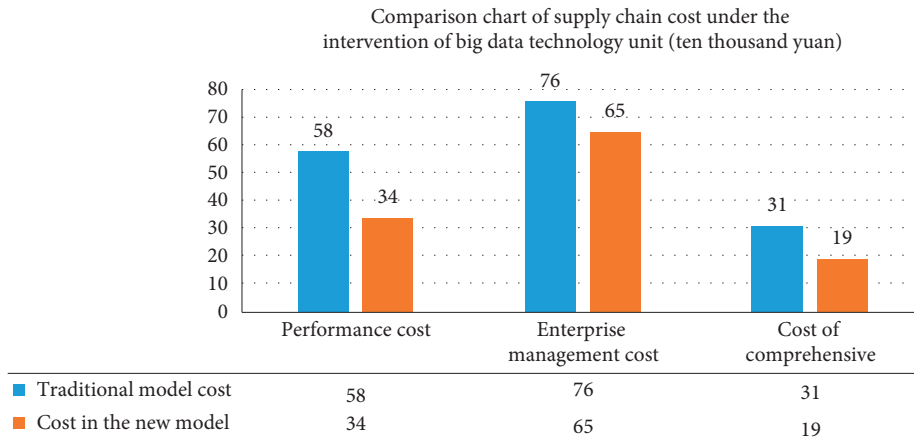


FIGURE 8: Comparison of SCCs under the intervention of big data.

TABLE 4: Running time when Co, GA, and PSO are applied to the dataset.

| Algorithm                   | 1   | 2   | 3    | 4    |
|-----------------------------|-----|-----|------|------|
| Collaborative optimization  | 850 | 800 | 4950 | 3760 |
| Genetic algorithm           | 650 | 580 | 3600 | 2660 |
| Particle swarm optimization | 320 | 200 | 2150 | 1770 |

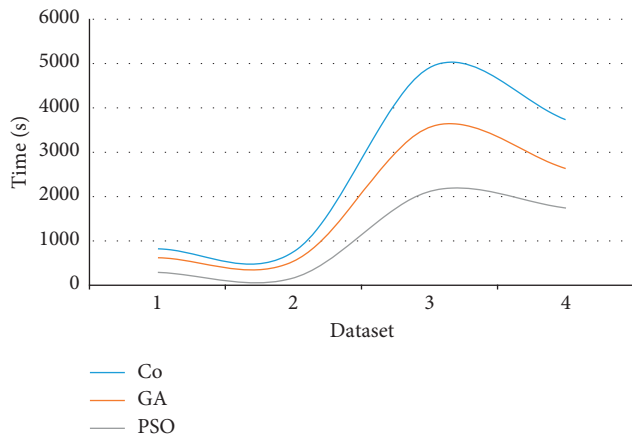


FIGURE 9: Comparison of Co, GA, and PSO when applied to the Plants dataset.

compared with the traditional model. The specific comparison figure is shown in Figure 8.

#### 4.4. Analysis of Data Mining Algorithms under Big Data.

The coevolutionary algorithm is implemented by Matlab7.6 programming under the Windows XP operating system, and the genetic algorithm, particle swarm algorithm, and coevolutionary algorithm are, respectively, applied to association rule mining by tracking the average fitness value and running time of the population during the evolution process. Performance pros and cons at the time.

Using genetic algorithm, particle swarm algorithm, and co-evolution algorithm for data mining and calculation, the Chess dataset, Solar Flare dataset, and the Plants dataset are obtained respectively, which is shown in Table 4. The dimensions of these two datasets are 6-dimensional and 10-dimensional, respectively, which belong to low-dimensional datasets. It can be seen from Figure 9 that Using the genetic algorithm, particle swarm algorithm, and co-evolution algorithm for data mining and calculation, the Chess dataset, Solar Flare dataset and the Plants dataset are obtained respectively. It can be seen from Figure 9 that in the early stage of evolution, the convergence speed and individual quality of the coevolutionary algorithm are slightly better than the other two algorithms. As the number of iterations continues to increase, the three algorithms all converge. From the experimental results, when mining association rules on low-dimensional datasets, the mining quality of the three is not very different, and the coevolutionary algorithm is slightly better than genetic algorithm and particle swarm algorithm.

## 5. Conclusion

This study mainly studies the trouble existing in the SCC command process of EE in the era of big data, proposes the management and command mode of SCC command based on big data technology, and analyzes the role that each supply chain link can play in this mode. In this study, the research for big data era background, promoting EE, adopts the strategy of big data view for cost control experience

theory instruction and guidance, gives full play to the author's knowledge and energy to help e-commerce businesses to get more profit by reducing cost, and vigorously promotes the further development of the electricity industry.

This study reflects the strong characteristics of new arguments in both the topic selection stage and the creation stage. First of all, this study discusses how to apply big data to EE to save the cost of supply chain under the background of booming big data technology. Second, in the creative process, this study changes the previous scholars' emphasis on theory and ignores the practical defects and focuses on the possibility of practice on the basis of summarizing the theory.

Through research, this study finds that there are still many improvements to be made in the practice of SCC command in the e-commerce industry. For example, most EE lack systematic analysis of costs, do not have contractual partners, serious waste of costs in sales, and wrong policies in logistics. Through the study of this research work, the conclusion is drawn that using big data technology for cost control can effectively solve many of the above problems, so as to continuously promote the progress of enterprise management and the reduction cost. The practical feature of this study lies in that it not only puts forward corresponding measures but also has a general significance for the CCE industry.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by special project of serving national strategy of Dalian Minzu University (2020fwgj056).

## References

- [1] C. Viola and S. Gammelgaard, "The influence of the digital divide on big data generation within supply chain management," *International Journal of Logistics Management*, vol. 29, no. 2, pp. 592–628, 2018.
- [2] T. Fan and Q. He, "A study of pricing and trading model of Blockchain & Big data-based Energy-Internet electricity," *Rapid Prototyping Journal*, vol. 19, no. 1, pp. 144–167, 2018.
- [3] J. Ruan, "An immune genetic algorithm for multi-echelon inventory cost control of IoT based supply chains," *IEEE Access*, vol. 323, no. 12, pp. 122–128, 2018.
- [4] S. Wang, "Cost optimization control of logistics service supply chain based on cloud genetic algorithm," *Wireless Personal Communications*, vol. 102, no. 4, pp. 3171–3186, 2018.
- [5] L. Han, "Cost-sharing contract of supply chain based on carbon emission control," *International Conference on Computational Intelligence & Security*, vol. 29, no. 1, pp. 56–68, 2018.

- [6] P. Cheng, "The effects of supply chain integration on the cost efficiency of contract manufacturing," *Journal of Supply Chain Management*, vol. 54, no. 3, pp. 34–87, 2018.
- [7] C.. Martin, "Regional manufacturing cost structures and supply chain considerations for sic power electronics in medium voltage motor drives," *Materials Science Forum*, vol. 924, no. 431, pp. 518–522, 2018.
- [8] C. Bolier, "Linking supply chain governance and biosecurity in the context control in western java: a value chain perspective," *Frontiers in Veterinary Science*, vol. 5, no. 2, pp. 94–121, 2018.
- [9] M. Roberto, "Pricing decision for information product supply chain under demand and cost disruption," *Computer Integrated Manufacturing Systems*, vol. 371, no. 264, pp. 62–67, 2018.
- [10] D. Ghosh and J. Shah, "Supply chain analysis under green sensitive consumer demand and cost sharing contract," *International Journal of Production Economics*, vol. 164, no. 4, pp. 319–329, 2015.
- [11] M. A. Jacobs, W. Yu, and R. Chavez, "The effect of internal communication and employee satisfaction on supply chain integration," *International Journal of Production Economics*, vol. 171, no. JAN., PT.1, pp. 60–70, 2016.
- [12] T. Supeekit, T. Somboonwivat, and D. Kritchanai, "Dematel-modified anp to evaluate internal hospital supply chain performance," *Computers & Industrial Engineering*, vol. 102, pp. 318–330, 2016.
- [13] D. Makepeace, P. Tatham, and Y. Wu, "Internal integration in humanitarian supply chain management," *Journal of Humanitarian Logistics and Supply Chain Management*, vol. 7, no. 1, pp. 26–56, 2017.
- [14] A. D. Smith, "Balancing internal supply chain logistics: a comparative analysis of manufacturing and service firm operations," *International Journal of Procurement Management*, vol. 3, no. 2, pp. 145–166, 2017.
- [15] J. M. Riley, R. Klein, J. Miller, and V. Sridharan, "How internal integration, information sharing, and training affect supply chain risk management capabilities," *International Journal of Physical Distribution & Logistics Management*, vol. 46, no. 10, pp. 953–980, 2016.
- [16] V. Turkulainen, J. Roh, J. M. Whipple, and M. Swink, "Managing internal supply chain integration: integration mechanisms and requirements," *Journal of Business Logistics*, vol. 38, no. 4, pp. 290–309, 2017.
- [17] S. Graham, "Antecedents to environmental supply chain strategies: the role of internal integration and environmental learning," *International Journal of Production Economics*, vol. 197, pp. 283–296, 2018.
- [18] Y. Agyabeng-Mensah, E. Ahenkorah, E. Afum, A. Nana Agyemang, C. Agnikpe, and F. Rogers, "Examining the influence of internal green supply chain practices, green human resource management and supply chain environmental cooperation on firm performance," *Supply Chain Management: An International Journal*, vol. 25, no. 5, pp. 585–599, 2020.
- [19] Kailash, R. K. Saha, and S. Goyal, "Benchmarking practice for identification of internal supply chain management performance factors gap," *Journal of Supply Chain Management Systems*, vol. 6, no. 4, pp. 33–38, 2017.
- [20] A. Agarwal, F. C. Giraud-Carrier, and Y. Li, "A mediation model of green supply chain management adoption: the role of internal impetus," *International Journal of Production Economics*, vol. 205, pp. 342–358, 2018.

## Research Article

# Complex System of Vertical Baduanjin Lifting Motion Sensing Recognition under the Background of Big Data

Yan Zhang,<sup>1</sup> M. M. Kamruzzaman ,<sup>2</sup> and Lu Feng <sup>3</sup>

<sup>1</sup>Sports Teaching and Research Office, Suzhou Institute of Trade and Commerce, Suzhou 215009, Jiangsu, China

<sup>2</sup>Department of Computer and Information Science, Jouf University, Sakaka 72311, Al-Jouf, Saudi Arabia

<sup>3</sup>Department of Basic Courses, Wuhan Donghu University, Wuhan 430212, Hubei, China

Correspondence should be addressed to Lu Feng; [lufeng@wdu.edu.cn](mailto:lufeng@wdu.edu.cn)

Received 23 October 2020; Revised 2 December 2020; Accepted 30 December 2020; Published 9 February 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Yan Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Nowadays, the development of big data is getting faster and faster, and the related research on motion sensing recognition and complex systems under the background of big data is gradually being valued. At present, there are relatively few related researches on vertical Baduanjin in the academic circles; research in this direction can make further breakthroughs in motion sensor recognition. In order to carry out related action recognition research on the lifting action of vertical Baduanjin, this paper uses sensor technology to collect the motion video image of vertical Baduanjin based on the background of big data and uses action recognition technology and related algorithms to obtain the action. Recognize the video image to obtain the data, get the acceleration, angular velocity, and EMG data, and count the end time and duration according to the change of the action. According to the data table and graph change trend compiled at the end of the experiment, we can see the following: after the data is preprocessed, the acceleration signal change range is limited to  $[-1, 1]$ , and the acceleration change has a clear directionality; and, after 15 lifts of the detected object, its angular velocity in  $X$ -axis direction is basically negative. However, when the ninth lift is performed, the angular velocity of the movement in  $X$ -axis direction is 36.09, the largest of all angular velocities. When performing the 15th lifting action, the angular velocity of this action in  $Z$ -axis direction is  $-26.05$ , which is the smallest of all angular velocities. The longest duration of the left muscle discharge during the lifting action of the subject is 15.24 for the tibial anterior muscle and 8.91 for the external oblique muscle with the shortest duration. The longest discharge duration of the right muscle is also the tibial anterior muscle with 12.15, and the shortest duration is the erector spinae with 8.79.

## 1. Introduction

**1.1. Background Meaning.** With the development of big data-related technologies, the applications of its scholars are increasingly being applied to big data human-computer interaction, motion recognition, and sensor technology. Nowadays, people are paying more and more attention to health-related sports, and the research on health and sports-related exercise recognition is also following closely. As an important part of healthy traditional culture, vertical Baduanjin has been bred in the soil of traditional Chinese culture for thousands of years. Its unique cultural tone and characteristics have aroused people's attention.

**1.2. Related Work.** Many scholars have done some researches on motion recognition. Wang proposed a new method, namely, weighted hierarchical depth motion map (whdmm) + three channel deep convolution neural networks (3convnets), which is used to recognize human actions from depth maps on small training data sets. Three strategies were developed to mine distinguishing features by using the ability of ConvNets [1]. Yan proposed an improvement to the soft attention model by combining convolutional long short-term memory (LSTM) with a layered system architecture to identify the action categories in the video, and Yan called this model a convolutional hierarchical attention model (CHAM), the model unit inside LSTM

convolution operation, and, using the focus map generation process to identify the action, layered architecture model is based on the particle size that can be more clearly the operation type [2].

*1.3. Innovation in This Article.* This paper studies the lifting motion of the vertical Baduanjin. The experimental research background is big data technology, using motion recognition related methods and sensor technology and using complex system theory to analyze the experiment. The innovations of this article are mainly reflected in the following aspects: (1) research on experiment in the context of big data, it is very convenient to collect data in experiments, which saves the time of data collection; (2) research on vertical Baduanjin lifting movements subject new and innovative matter; (3) the analysis of experiments using complex systems theory can draw more scientific and reasonable experimental results.

## 2. Relevant Technology of Vertical Baduanjin Lifting Motion Sensing Recognition

*2.1. Big Data.* As an important factor in production, data has penetrated into every industry today, and the efficiency of data mining and implementation will directly affect the development of the new productivity cycle. Big data is a huge, diverse, and rapidly changing information asset. It needs to adopt new processing methods to improve people's ability in vision expansion, process optimization, and decision formation [3, 4]. The volume of big data greatly exceeds the capabilities of commonly used software tools, and it is difficult to obtain, manage, grasp, and process within a tolerable time. There are many specific processing methods for data, but the current general process of big data processing can be summarized into four steps: big data collection, big data input and preprocessing, big data analysis and statistics, and big data mining [5]. Its specific processing flow is shown in Figure 1.

Big data analysis process is collected and introduced into the first structured, semistructured, and unstructured data and then converts the data quality and loading operation by the data integration. If the data needs to be processed in real time, data will then flow into the computing module; if the data does not need to be processed in real time, the data can first enter the data integration module data quality screening; filtration pretreatment may not be direct via the data integration module and offline processing incoming data analysis module. It is necessary to conduct modeling and predictive analysis by mining data analysis tools. Finally, when data reaches the presentation layer, management can make decisions, discovery, modeling and forecasting analysis, reporting, planning prophecy, and content analysis applications show. The operation of large data sharing and data sharing module refers to a repeating operation in different applications or data information may be precipitated into the module, the data for later analysis of other common applications. The analysis and processing process of big data can be represented by a big data analysis architecture diagram, which is shown in Figure 2.

Large and complex data only exist as data material before being identified and discovered. When people have discovered its value but have not yet begun to classify its value they can only be regarded as a data resource [6]. Only after collection and merging, valuable information can be extracted from data analysis tools to form data assets and utilize resources [7]. Although the data materials, data resources, and data assets in big data have the same material connotation, they belong to different management categories and have significant differences. The relationship between the three is shown in Figure 3.

The characteristics of big data can be summarized as four V: volume, velocity, variety, and value. The amount of data aggregated by big data is very huge, and the unit of measurement has usually reached PB (1 million G), EB (1 billion G), and ZB (1 trillion G). The speed of big data is reflected in two aspects: first, the speed of data generation is fast. The other is the need for very fast data processing speed. Being fast is the biggest difference between big data technology and traditional data technology. In the era of big data, not only the amount of data has surged, but also the sources and types of data have become more and more complex. The huge amount of traditional data is usually predefined structured data. With the emergence and development of detection technology and the Internet, semistructured and unstructured data have emerged in large numbers, enriching data. Diversity also increases the difficulty of processing and storing data. However, current data usually has the problem of low value density. As the data grows, the amount of useful information that can be extracted from the massive data may be very small and cannot increase according to the amount of data. It is more obvious in unstructured and semistructured data. Therefore, how to use powerful data calculation methods to quickly obtain data value and improve the quality of data information is one of the urgent problems to be solved in the context of the current rapid development of big data.

*2.2. Sensor Technology.* A sensor is a device that can sense the measured object and convert it into a signal that can be used to transmit a signal according to a specific rule. It is also an energy conversion device that can convert one energy into another energy [8, 9]. The sensor is mainly composed of two parts: the sensitive element and the conversion element. Some sensors also include signal amplification circuits, mainly because the electrical signal output from the sensor is relatively weak and difficult to measure [10]. The block diagram of the sensor is shown in Figure 4.

The relationship between the sensor and the detection technology is very close. The detection technology is the window through which the sensor obtains the measurement. The sensor is at the interface of the object to be measured and the detection system. You need to understand the detection technology before you can better use the sensor [11]. The most important thing in detection technology is the measurement of the detected object. The measurement can be expressed by the following formula:

$$x = nu, \quad (1)$$

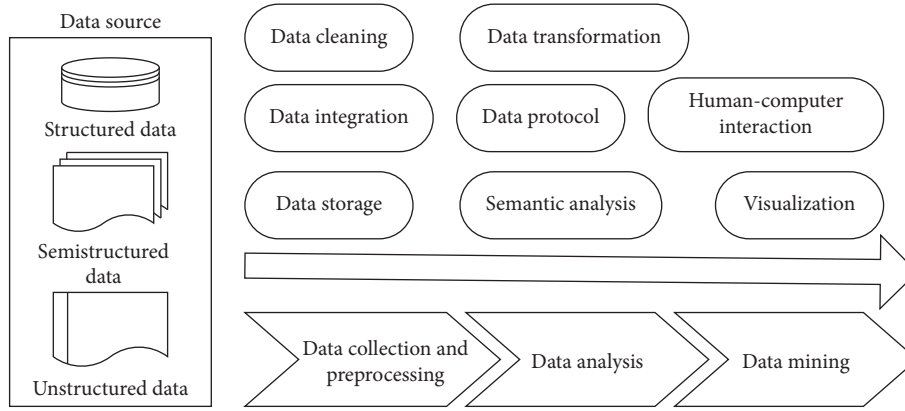


FIGURE 1: Big data processing flow.

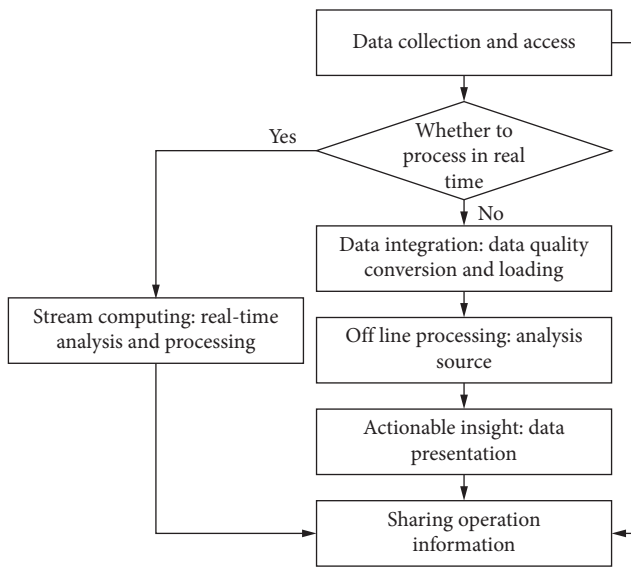


FIGURE 2: Big data analysis architecture diagram.

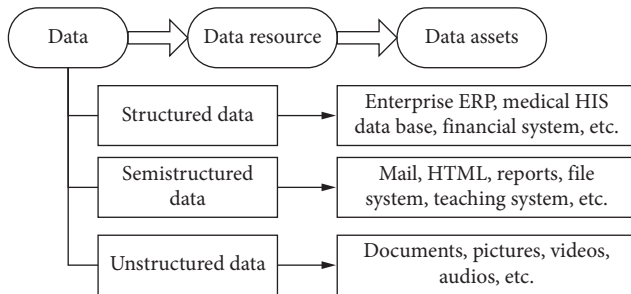


FIGURE 3: Relationship diagram of data materials, data resources, and data assets.

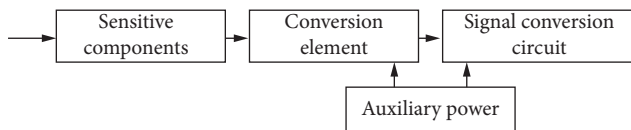


FIGURE 4: Sensor composition block diagram.

where  $x$  is the measured value,  $u$  is the standard quantity, the unit of measurement, and  $n$  is the multiple of the measurement object. The direct detection method of detection technology can be expressed by the first following formula and the indirect detection method can be expressed by the second following formula:

$$y = x, \tag{2}$$

$$y = f(x). \tag{3}$$

In the detection system, the sensor is connected to the measured object and the signal conditioning circuit. The signal conditioning circuit acquires the measured object through the sensor and then processes the signal. The basic framework of the detection system is shown in Figure 5.

In the open-loop measurement of the detection system, the direction of information transition is in one direction. There are two channels for the information direction in the closed-loop measurement, one is the positive channel, and the other is the feedback channel. Their input and output relationship formulas are shown in the following formulas respectively:

$$y = k_1 k_2 k_3 x, \tag{4}$$

$$y = \frac{k k_1}{1 + k \beta} x \approx \frac{k_1}{\beta} x, \tag{5}$$

where  $k_1, k_2, k_3$  is the transfer coefficient between each link,  $\beta$  is the feedback coefficient of closed-loop measurement, and  $k = k_2 k_3$ . In the actual measurement process, due to the influence of environmental factors and human factors, it is difficult to achieve accurate measurement, which will cause measurement errors. The expression methods of measurement error mainly include absolute error, actual relative error, quoted error, basic error, and additional error. The formula for absolute error is as follows:

$$\Delta = x - L. \tag{6}$$

The actual relative error formula is as follows:



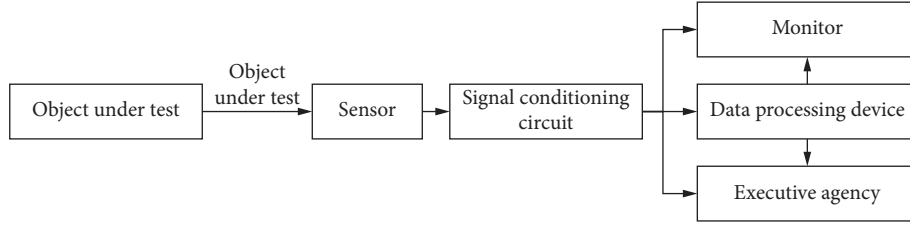


FIGURE 5: The basic framework of the detection system.

$$\delta = \frac{\Delta}{L} \times 100\%. \quad (7)$$

The formula for citation error is as follows:

$$\gamma = \frac{\Delta}{\max - \min} \times 100\%, \quad (8)$$

where  $\Delta$  represents absolute error,  $x$  is the measured value,  $L$  is the true value, and  $\max$ ,  $\min$  are the true value.

The relationship between sensor input and output is a key feature of the sensor. Different sensors produce different forms of electrical energy, which mainly depends on the basic characteristics of the sensor. The sensor has static characteristics and dynamic characteristics. If the input signal is static, then the relationship between the output  $y$  of the sensor and the input  $x$  does not change with time; that is, the output can be expressed by an equation without time:

$$y = a_0 + a_1x + a_2x^2 + \dots + a_nx^n, \quad (9)$$

where  $a_0$  represents the output when the input is equal to 0 and  $a_1, a_2, \dots, a_n$  is the nonlinear coefficient. The indicators of the static characteristics of the sensor include the sensitivity, linearity, repeat-ability, hysteresis, and drift of the sensor. Their formula expression is as follows:

$$\text{sensitivity: } S = \frac{\Delta y}{\Delta x},$$

$$\text{linearity: } \gamma_L = \pm \frac{\Delta L_{\max}}{Y_{FS}} \times 100\%, \quad (10)$$

$$\text{hysteresis: } \gamma_H = \frac{\Delta H_{\max}}{Y_{FS}} \times 100\%,$$

$$\text{temperature drift: } W = \frac{y_t - y_{20}}{Y_{FS} * \Delta t} \times 100\%.$$

$S$  in the formula represents the sensitivity of the sensor,  $\Delta L_{\max}$  is the absolute error of the maximum nonlinearity,  $Y_{FS}$  is the output value of the full scale of the sensor,  $\Delta H_{\max}$  is the hysteresis difference of the output,  $y_t$  is the output output temperature, and  $y_{20}$  is the output at temperature 20 degrees Celsius when the output, and  $\Delta t$  is the difference between the temperature  $t$  and 20 degrees Celsius.

The dynamic characteristics of the sensor mean that the output of the sensor changes with time; that is, the output of the sensor is a time function of the input. For ideal dynamic characteristics, the output time function and the input time function are not the same, and the difference between them

lies in the dynamic error. The dynamic equation of the sensor is as follows:

$$\begin{aligned} a_n \frac{d^n y}{dt^n} + a_{n-1} \frac{d^{n-1} y}{dt^{n-1}} + \dots + a_1 \frac{dy}{dt} \\ + a_0 y = b_m \frac{d^m y}{dt^m} + b_{m-1} \frac{d^{m-1} y}{dt^{m-1}} \\ + \dots + b_1 \frac{dy}{dt} + b_0 x. \end{aligned} \quad (11)$$

**2.3. Action Recognition.** The process by which the computer can analyze and judge various behaviors and movements through automatic detection is called action recognition [12]. The main task is to analyze the motion recognition image or video collection to determine the sequential images or motion video, using a video editing operation of the computer image, using a computer to locate the last track, the video image to identify a target. The traditional action recognition method is to manually extract the graphic features from the video or image, then reduce the dimensionality of the generated feature vector, and finally use different classifiers for classification and recognition [13]. According to the description method used in the video recognition process, action recognition technology can be roughly divided into the following four categories: Patios-temporal feature recognition method, probability recognition and reasoning method, action rule recognition method, and action pattern recognition method. According to the number of levels of action features, action recognition technology can be divided into single-level recognition methods and multilevel recognition methods. It is widely used in video surveillance, human-computer intelligent interaction, medical care, virtual reality, sports, and other fields. It is an interesting and challenging problem. The application areas of action recognition in young and old people are shown in Figure 6.

Action recognition can obtain raw data by using video equipment and wearable equipment according to the difference of sensor electronic equipment [14]. The motion recognition method using video equipment is a computer vision-based system method, and the motion recognition method using a wearable device is a sensor-based system method. With the increasing maturity of the manufacturing process of microelectronic mechanical systems, various electronic components and products, especially sensors, have been greatly improved and developed rapidly in terms

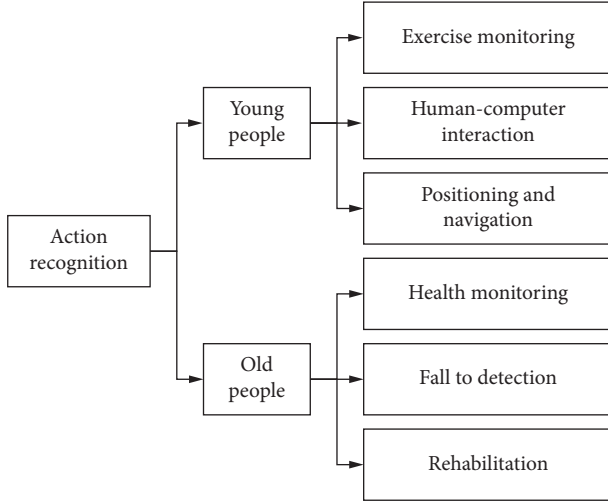


FIGURE 6: Action recognition application area diagram.

of volume, accuracy, power consumption, and performance. Sensors play an increasingly important role in action recognition [15, 16]. There are many ways to recognize actions; the basic idea is to match the template to be classified with the template of the known category. By measuring the similarity between the patterns, the unknown category can be classified. The second is a generative model-based method that uses a common probability function to determine the relationship between the observed attribute value and the action category information. Another method is based on the recognition model, which can directly model the conditional probability of a given category of operators. The classification model of this method is relatively simple and the calculation efficiency is high.

**2.4. Complex System.** Complex systems are the main research objects of complexity science and can be found in every corner of daily life. Life system and social system are complex systems [17]. Complex systems are usually composed of multiple interactive units, which have characteristics or features that not every component unit has. For a complex system, its total performance is not a simple superposition of partial performance, and the relationship between total performance and partial performance is a complex nonlinear relationship [18]. The complex system emphasizes the hierarchical relationship and appearance characteristics between the individual and the whole and pays attention to dynamic issues in the overall evolution of the system. The interaction of various factors in the system, various subsystems, and the environment in which the system is located makes the system continuously upgrade and develop. It does not depend on one or a specific pair of influencing factors, and the system is very sensitive to this interaction. Small changes between projects can lead to completely different results [19, 20]. Complex systems have the characteristics of openness, hierarchical structure, complexity, dynamics, and nonlinearity.

The complex system is still in its infancy and may include a new system or even a revolution in traditional scientific

methods. The complex system is constantly developing and evolving under the action of force. The dynamic process of the entire complex system can be described by the dynamic behavior of the individual. The complexity and interdisciplinary nature of the complex system field determines that there is no proprietary model in the complex system field [21, 22]. The basic model of a complex system is a cellular automaton, which is a dynamic system. Research on the complex evolution of cellular automation system behavior and its occurrence mechanism is helpful to explore the complexity of complex systems and the study of their internal laws [23, 24]. Cellular automata have different times, different states, different spaces, homogeneity, normal positions, and parallelism.

Cellular automata can be represented by a four-tuple,  $S$  is the state of the cell,  $L_d$  is the discrete cell space,  $N$  is the cell neighborhood of finite radius, and  $f$  is the local rule:

$$A = (S, L_d, N, f). \quad (12)$$

The expression of elementary cellular automata: assuming that the number of cell states is  $k$ , its state set is  $\{s_1, s_2, \dots, s_k\}$ , the neighbor radius is  $r$ , and the cell state at time  $t$  is  $s_i^t$ ; the neighbor cell of this cell and itself is called neighborhood cells, and its state is

$$s = \{s_{i-r}^t, \dots, s_{i-2}^t, s_{i-1}^t, s_i^t, s_{i+1}^t, s_{i+2}^t, \dots, s_{i+r}^t\}. \quad (13)$$

An arrangement of state values of neighboring cells will correspond to a regular function  $f$ . Suppose that the state of the cell at the next moment is  $s_i^{t+1}$  and  $f$  is defined as follows:

$$s_i^{t+1} = f(s). \quad (14)$$

The state set of elementary cellular automata is usually defined as  $\{0, 1\}$ , and its state transition function can be expressed as

$$s_j^{t+1} = f(s_{j-1}^t, s_j^t, s_{j+1}^t). \quad (15)$$

The cell state at time  $t$  is  $s_j^t$ .  $D$ ,  $s_{j-1}^t, s_{j+1}^t$ , are the states of the left neighbor and right neighbor of the cell, respectively.

Reversible cellular automata can be used in information security fields such as information encoding, image encryption, and decryption, and its formula is described as

$$s_i^{t+1} = f(s_{i-r}^t, \dots, s_{i-2}^t, s_{i-1}^t, s_i^t, s_i^{t-1}, s_{i+1}^t, s_{i+2}^t, \dots, s_{i+r}^t). \quad (16)$$

All the cells of a two-dimensional cellular automaton are distributed in a plane grid composed of regular cells, and its cell state at the next moment can be expressed as

$$s_{i,j}^{t+1} = f(s_{i-1,j}^t, s_{i+1,j}^t, s_{i,j-1}^t, s_{i,j+1}^t). \quad (17)$$

### 3. Vertical Baduanjin Lifting Motion Sensor Recognition Experiment Design

**3.1. Data Collection.** The recognition of lifting motion is divided into two parts: data collection end and data processing end. Sensor-based action recognition is a process of

data collection and feature extraction and classification. The data collection terminal can be regarded as a whole, which includes the main control module, sensor module, and communication module. The block diagram of the data acquisition terminal is shown in Figure 7.

The data collection of actions can be divided into three types according to the data transmission method: the use of wired connection, the wireless transmission, and the direct reading of the data through the local data storage method. The wireless transmission of data collected can only increase the comfort performance data collection system that is worn and can observe real-time data acquisition storage; thus, it is superior choice. This experiment uses sensors as the transmission channel for wireless data transmission, uses sensor technology to identify motion data, and transmits the motion data to the receiving end. The processing flow of data collection is shown in Figure 8.

**3.2. Action Data Preprocessing.** In the process of collecting experimental data, even the best data acquisition equipment will inevitably introduce noise into the experimental data stream. The appearance of noise will affect the follow-up research work and interfere with the judgment of the final experimental results. This experiment uses smoothing denoising and normalized data preprocessing methods to denoise the acquisition process of the acceleration signal and angular velocity signal of the upper limb lifting motion of the measured object. Many studies of denoising embodiment implemented using wavelet denoising noisy signals to the original signal select a wavelet basis function and decomposition level  $n$  (generally  $n$  takes 3–5) for wavelet decomposition, select a threshold to quantize the high-frequency coefficients of each level wavelet decomposition, and then reconstruct the low-frequency coefficients and high-frequency coefficients after threshold quantization. The flowchart of wavelet threshold denoising is shown in Figure 9.

There are mainly two commonly used threshold functions: hard threshold function and soft threshold function. Suppose that  $\omega$  is the original wavelet coefficient and  $T$  is the set threshold; then, their expressions are

$$\eta(\omega) = \begin{cases} \omega, & |\omega| \geq T, \\ 0, & |\omega| < T, \end{cases} \quad (18)$$

$$\eta(\omega) = \begin{cases} \text{sgn}(\omega)(|\omega| - T), & |\omega| \geq T, \\ 0, & |\omega| < T. \end{cases}$$

Due to the differences between individuals, different people perform the same upper limb movement; their acceleration performance, including movement amplitude and execution speed, is always different. Through the normalization technology, the user's action acceleration data of different intensities (the amplitude of the acceleration signal) can be adjusted. This experiment uses normalization to standardize the amplitude range of the original acceleration data to the same amplitude range of  $[-1, 1]$ . Assuming that the original data is  $P = \{p_1, p_2, \dots, p_n\}$ , normalize it to  $[m, n]$  above the interval, and the normalized data is

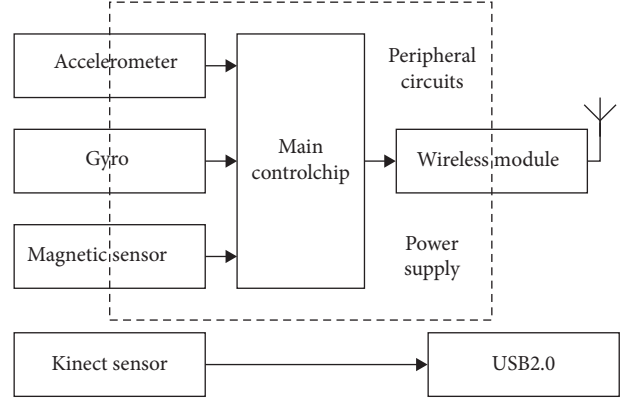


FIGURE 7: Block diagram of the data acquisition terminal.

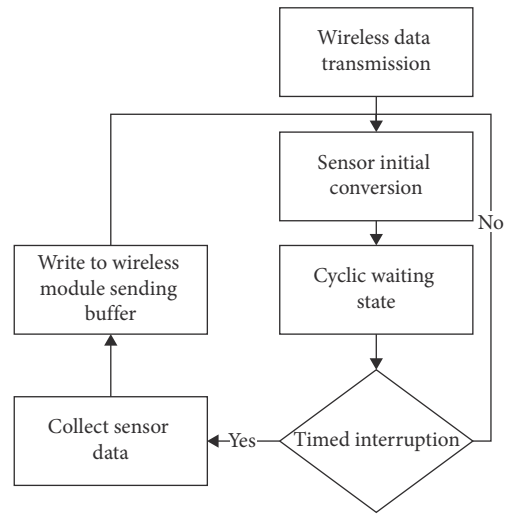


FIGURE 8: Data acquisition processing flowchart.

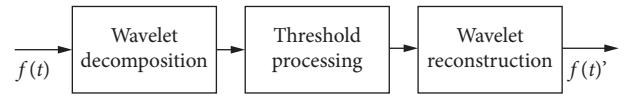


FIGURE 9: Wavelet threshold denoising flowchart.

represented by  $T = [t_1, t_2, \dots, t_n]$ ; then, the normalized formula is as follows:

$$t_i = m + \frac{(n - m) * (p_i - \min(p))}{(\max(p) - \min(p))}. \quad (19)$$

**3.3. Feature Parameter Extraction.** The time-frequency analysis method is used to extract the characteristic parameters of the action, which is based on the wavelet theory. Denote the wavelet scaling function  $\Phi(t)$  and the wavelet function  $\Psi(t)$  by  $\Phi(t) = u_0(t)$  and  $\Psi(t) = u_1(t)$ , respectively, and use  $u_n(t)$  to satisfy the following two-scale equation:

$$\begin{cases} u_{2n}(t) = \sqrt{2} \sum_{k \in \mathbb{Z}} h_k u_n(2t - k), \\ u_{2n+1}(t) = \sqrt{2} \sum_{k \in \mathbb{Z}} g_k u_n(2t - k). \end{cases} \quad (20)$$

The conjugate filter coefficient  $g_k = (-1)^k h_{(1-k)}$  in the formula has an orthogonal relationship. The  $\{u_n(t)\} (n \in Z^+)$  constructed by formula (20) is called the wavelet packet determined by the basis function  $\Phi(t) = u_0(t)$ . The wavelet packet of the signal is expressed as

$$g_j^n(t) = \sum_{k \in Z} d_l^{j,n} u_n(2^j t - k). \quad (21)$$

The collected three-dimensional acceleration information of the upper limbs of the human body has signals in X, Y, and Z directions that are mutually perpendicular. For each dimensional acceleration signal, the same wavelet packet decomposition is performed to obtain the energy of each frequency band in three directions.

#### 4. Data Analysis of Motion Sensing Recognition of Vertical Baduanjin Lifting

**4.1. Acceleration Signal Comparison before and after Motion Data Preprocessing.** The normalized motion acceleration signals have the same maximum value, so the motion acceleration signals of different test objects are integrated according to the same standard, thereby creating conditions for feature parameter extraction and motion classification and recognition. In this experiment, after preprocessing the relevant data of the vertical Baduanjin lifting movement, the acceleration signal changes of this movement are shown in Table 1.

According to the data in Table 1, we can see the change of acceleration signal before and after the data preprocessing of the action. In order to observe the change trend of the acceleration signal more intuitively, we compare the change of the acceleration signal of the action. We convert the data in Table 1 into a graphic way, and the final conversion result is shown in Figure 10.

According to the graph change in Figure 10, we can see that the acceleration variation range without pretreatment is between  $[-1500, 1000]$  and the data change span is large. After preprocessing, the amplitude of acceleration signal change is limited between  $[-1, 1]$ , and the change of acceleration has clear direction.

**4.2. Angular Velocity Change of Lift.** After the feature extraction of the detected object, the angular velocity signal changes of the object in 15 lifts are counted, and the angular velocity changes in X, Y, and Z directions are analyzed. According to the data acquisition terminal, the angular velocity data of 15 times of lifting actions are shown in Table 2.

According to the data in Table 2, we can understand the angular velocity change of the action after 15 lifts. In order to change the angular velocity of the operation for visual analysis, we converted the data in Table 2 and the final result is shown in Figure 11.

According to the data in Figure 11, we can see that, after 15 lifts, the angular velocity of the action is basically negative in x-axis direction; however, in the ninth lift, the angular velocity of the action in x-axis direction is 36.09, which is the

TABLE 1: Action acceleration signal change table.

| Sampling times | before |      |      | after |       |       |
|----------------|--------|------|------|-------|-------|-------|
|                | X      | Y    | Z    | X     | Y     | Z     |
| 0              | -100   | 900  | 0    | 0.95  | 0.92  | 1     |
| 30             | -120   | 950  | 20   | 1     | 1     | 1     |
| 60             | -1200  | 0    | -400 | -1    | -0.6  | -0.8  |
| 90             | -600   | -400 | -625 | -0.6  | -0.9  | -1    |
| 120            | -800   | -450 | -500 | -0.45 | -1    | -0.7  |
| 150            | -850   | -455 | -520 | -0.55 | -0.95 | -0.85 |

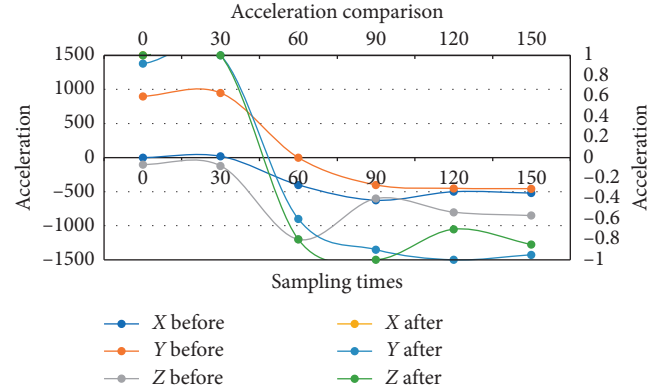


FIGURE 10: Comparison of acceleration signal before and after data preprocessing.

largest among all the angular velocities; in the 15th lifting action, the angular velocity of the action in z-axis direction is  $-26.05$ , which is the smallest among all the angular velocities.

**4.3. Muscle EMG Test Results of Lifting Motion.** The electrocardiography signals of tibial anterior muscles (TAM), erector spinae (ES), rectus abdominis (RA), external oblique muscles (EOM), and gluteus maximus (GM) were collected, and the trigger time, end time, and duration were analyzed. The statistical data of muscle electricity are shown in Table 3.

According to the data in Table 3, we can know the trigger time, end time, and duration of the EMG signal of some muscles when the body is performing the vertical Baduanjin lifting action. Convert the data in Table 3 to the left side of the body. The EMG signal time data graph of the side muscles and the EMG signal time data graph of the right muscles of the body: the converted results are shown in Figures 12 and 13.

According to the data in Figures 12 and 13, we can see the discharge sequence and sustained discharge time of part of the muscles of the test object during the lifting action. From Figure 12, we can see that the longest duration of the left muscle discharge is the tibial anterior muscle at 15.24 and the shortest duration is the external oblique muscle at 8.91. From Figure 13, we can see that the longest duration of the right muscle discharge is also the tibial anterior muscle with 12.15 and the shortest duration is the erector spinae with 8.79.

TABLE 2: Action angular velocity change table.

| Sampling point | X      | Y      | Z      |
|----------------|--------|--------|--------|
| 1              | -7.53  | -25.62 | 25.28  |
| 2              | -20.31 | -7.98  | 13.77  |
| 3              | -17.79 | 16.15  | 10.12  |
| 4              | -8.66  | -30.45 | 5.16   |
| 5              | -25.63 | 2.17   | -2.19  |
| 6              | 4.32   | -12.53 | -8.67  |
| 7              | -3.25  | -4.33  | -13.77 |
| 8              | -0.94  | -10.04 | -15.92 |
| 9              | 36.09  | -12.35 | -4.24  |
| 10             | -14.51 | 17.98  | 8.57   |
| 11             | -25.46 | 1.82   | 12.71  |
| 12             | -15.48 | 2.39   | 3.22   |
| 13             | -1.12  | -23.74 | -11.08 |
| 14             | 11.19  | 2.67   | -26.05 |
| 15             | 4.31   | -3.96  | -13.24 |

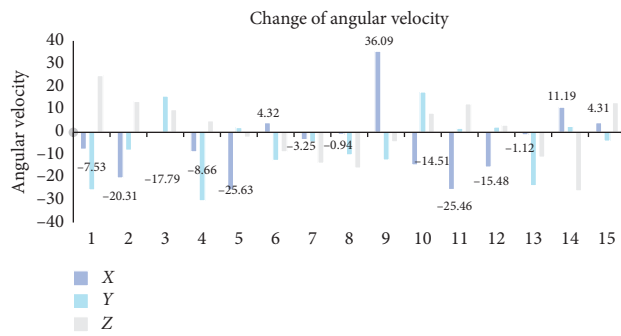


FIGURE 11: Action angular velocity change graph.

TABLE 3: Timetable of changes in muscle power.

| Muscle | Trigger time | Left     |          |              | Right    |          |  |
|--------|--------------|----------|----------|--------------|----------|----------|--|
|        |              | End time | Duration | Trigger time | End time | Duration |  |
| TAM    | -0.48        | 13.98    | 15.24    | 2.14         | 12.69    | 12.15    |  |
| ES     | -1.06        | 14.33    | 15.14    | 2.91         | 10.11    | 8.79     |  |
| RA     | 0            | 13.14    | 13.14    | 1.08         | 11.11    | 10.25    |  |
| EOM    | 1.23         | 11.04    | 8.91     | 0.5          | 12.14    | 10.81    |  |
| GM     | 1.04         | 14.32    | 13.01    | 1.17         | 9.95     | 9.53     |  |

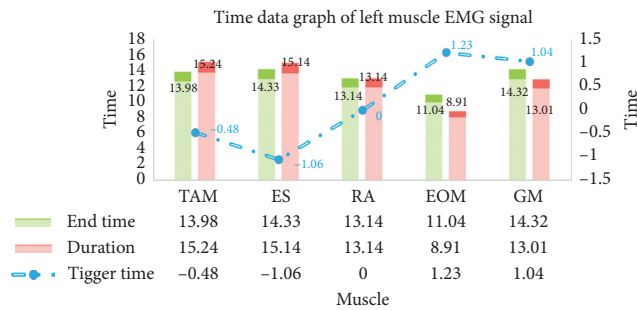


FIGURE 12: Time data graph of left muscle EMG signal.

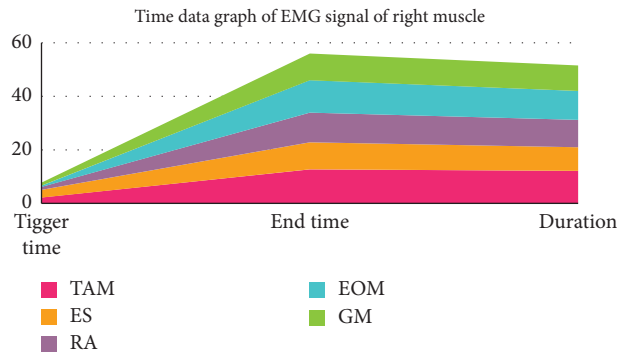


FIGURE 13: Time data graph of EMG signal of right muscle.

## 5. Conclusions

Big data is an emerging technology closely related to data collection, data transmission, and data analysis in recent years. It can also be the sum of a type of complex and huge data. Experimental research in the context of big data can solve the problem of experimental data collection. Combining big data, sensor technology, and complex system theory effectively and conducting related research in the direction of motion recognition will help scholars find technological breakthroughs and innovations in the field of motion recognition.

This research uses big data as the background to design experiments for the motion recognition of the vertical Baduanjin lifting motion, using sensor technology to collect the video image of the vertical Baduanjin lifting motion of the detected object, and use motion recognition related methods and formulas to compare motion video data for calculation and finishing; the final tally triggers time action acceleration, angular velocity, and part of the muscle EMG signal detection object, end time, and duration of the use of complex systems theory to experimental data analysis.

Although this experiment was performed well, there are still some shortcomings: first, although big data solves the problem of data collection well, the data in it is too large and difficult to store. Secondly, after the action video image was collected, the video image was not optimized, which caused some errors in the experimental data. Finally, the complex system theory is not very suitable for the research of vertical Baduanjin lifting motion sensing recognition, and it should be improved.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by Jouf University, Sakaka, Al-Jouf, Saudi Arabia.

## References

- [1] P. Wang, W. Li, Z. Gao, J. Zhang, C. Tang, and P. O. Ogunbona, "Action recognition from depth maps using deep convolutional neural networks," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 4, pp. 498–509, 2016.
- [2] S. Yan, J. S. Smith, W. Lu et al., "CHAM: action recognition using convolutional hierarchical attention model," *IEEE Transactions on Human-Machine Systems*, vol. 1, no. 1, pp. 3958–3962, 2017.
- [3] S. Athey, "Beyond prediction: using big data for policy problems," *Science*, vol. 355, no. 6324, pp. 483–485, 2017.
- [4] A. D. Mauro, M. Greco, and M. Grimaldi, "A formal definition of Big Data based on its essential features," *Library Review*, vol. 65, no. 3, pp. 122–135, 2016.
- [5] M. M. Rathore, A. Ahmad, A. Paul, and S. Rho, "Urban planning and building smart cities based on the Internet of Things using Big Data analytics," *Computer Networks*, vol. 101, no. C, pp. 63–80, 2016.
- [6] Z. Su, Q. Xu, and Q. Qi, "Big data in mobile social networks: a QoE-oriented framework," *IEEE Network*, vol. 30, no. 1, pp. 52–57, 2016.
- [7] L. Xu, C. Jiang, J. Wang et al., "Information security in big data: privacy and data mining[J]," *IEEE Access*, vol. 2, no. 2, pp. 1149–1176, 2017.
- [8] F. J. Blaauw, H. M. Schenk, B. F. Jeronimus et al., "Let's get Physical-an intuitive and generic method to combine sensor technology with ecological momentary assessments," *Journal of Biomedical Informatics*, vol. 63, no. 1, pp. 141–149, 2016.
- [9] K. Harada, J. Ishida et al., "Introduction to the special issue on "State-of-the-art sensor technology in Japan 2012," *Sensors*, vol. 14, pp. 11045–11048, 2016.
- [10] D. Wagner, S. Vogt, F. I. Jamal et al., "Application of microwave sensor technology in cardiovascular disease for plaque detection," *Current Directions in Biomedical Engineering*, vol. 2, no. 1, pp. 273–277, 2016.
- [11] O. S. Albahri, A. S. Albahri, K. I. Mohammed et al., "Systematic review of real-time Remote health monitoring system in triage and priority-based sensor technology: taxonomy, open challenges, motivation and recommendations," *Journal of Medical Systems*, vol. 42, no. 5, pp. 1–27, 2018.
- [12] S. He, C. Wu, and J. Tian, "Overview of current sensor technology," *Electrical Drive*, vol. 48, no. 1, pp. 65–75, 2018.
- [13] W. Du, Y. Wang, and Y. Qiao, "Recurrent spatial-temporal attention network for action recognition in videos," *IEEE Transactions on Image Processing*, vol. 27, no. 99, pp. 1347–1360, 2017.

- [14] N. N. Bui, J. Y. Kim, and H.-G. Kim, "Gradient-flow tensor divergence feature for human action recognition," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 99, no. 1, pp. 437–440, 2016.
- [15] S. Chun and C. S. Lee, "Human action recognition using histogram of motion intensity and direction from multiple views," *Iet Computer Vision*, vol. 10, no. 4, pp. 250–257, 2016.
- [16] D. Alyssa, "Temperature transmitter now includes Associated Intrinsically-Safe sensor connections," *Manufacturing Automation: Machine Design, Systems, Technology*, vol. 33, no. 1, p. 29, 2018.
- [17] D. G. Dessavre, J. E. Ramirez-Marquez, and K. Barker, "Multidimensional approach to complex system resilience analysis. Reliability Engineering," *System Safety*, vol. 149, pp. 34–43, 2016.
- [18] S. Lehuta, R. Girardin, S. Mahévas, M. Travers-Trolet, and Y. Vermard, "Reconciling complex system models and fisheries advice: practical examples and leads," *Aquatic Living Resources*, vol. 29, no. 2, pp. 208–209, 2016.
- [19] J. Chen, B. M. Chen, and J. Sun, "Complex system and intelligent control: theories and applications," *Frontiers of Information Technology & Electronic Engineering*, vol. 20, no. 1, pp. 1–3, 2019.
- [20] A. Almalaq, J. Hao, J. J. Zhang, and F.-Y. Wang, "Parallel building: a complex system approach for smart building energy management," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 6, pp. 1452–1461, 2019.
- [21] M. R. Pourhassan, S. Raissi, and A. Hafezalkotob, "A simulation approach on reliability assessment of complex system subject to stochastic degradation and random shock," *Eksploatacja I Niezawodnosc - Maintenance and Reliability*, vol. 22, no. 2, pp. 370–379, 2020.
- [22] J. Li and Y. Wang, "Coupling effect of regional industrial cluster and innovation based on complex system metric and fuzzy mathematics," *Journal of Intelligent & Fuzzy Systems*, vol. 37, no. 5, pp. 6115–6126, 2019.
- [23] Z. Xiong and T. Yao, "Service-oriented model encapsulation and selection method for complex system simulation based on cloud architecture," *Entropy*, vol. 21, no. 9, pp. 891–892, 2019.
- [24] B. Dadvand and F. Behzadpoor, "Pedagogical knowledge in English language teaching: a lifelong-learning, complex-system perspective," *London Review of Education*, vol. 18, no. 1, pp. 107–126, 2020.

## Research Article

# Earthquake Disaster Rescue Model Based on Complex Adaptive System Theory

Fujiang Chen <sup>1</sup>, Jingang Liu,<sup>2</sup> and Junying Chen<sup>2</sup>

<sup>1</sup>School of Emergency Science, Xihua University, Chengdu 610039, Sichuan, China

<sup>2</sup>China Railway Academy Co., Ltd., Chengdu 610000, Sichuan, China

Correspondence should be addressed to Fujiang Chen; 1220190024@mail.xhu.edu.cn

Received 23 October 2020; Revised 22 December 2020; Accepted 21 January 2021; Published 5 February 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Fujiang Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

China is located in the intersection area of two seismic zones. Due to this special geographical location, earthquake disasters occur frequently in China. Earthquake emergency rescue work is one of the key construction works of disaster prevention and mitigation in China. This paper mainly studies the earthquake disaster rescue model based on the complex adaptive system theory and establishes the earthquake disaster rescue model by analyzing the complex adaptive system theory and combining the earthquake rescue process. In this paper, through the task allocation mechanism task, the disaster rescue task is divided into simple task and complex task, and the executive task subject is divided into single task subject and multitask subject. On the basis of considering the shortest emergency rescue time goal and the goal of maximizing the deployment utility of rescue team, the reasonable deployment of a rescue team is realized through a complex adaptive system, that is, the deployment utility of the rescue team is maximized. In this paper, the simulation experiment and comparison of the earthquake disaster rescue model based on the complex adaptive system theory are carried out. The experimental results show that the model used in this paper is better than the other two models in terms of algorithm convergence, rescue number, and overall score; in different scenarios, the relative survival probability of the model in this paper is 58.92%, 67.85%, and 77.46%, and the proportion of the wounded rescued is 66.31%, 76.45%, and 83.06%, which were higher than those of the other two models. The earthquake disaster rescue model based on the complex adaptive system theory proposed in this paper provides an effective theoretical basis and method system for postdisaster emergency rescue decision making and enhances and improves the emergency response ability to deal with large-scale geological disaster events.

## 1. Introduction

Large-scale earthquake disaster has become the focus and difficulty of research because of its wide influence range, huge population, serious economic loss, high uncertainty, derivation, and evolution. In recent years, large-scale earthquake disasters occur frequently in the world, which is a serious threat to the safety of human life and property. For example, the Wenchuan earthquake caused more than 80000 deaths and missing people and 370000 injured people. A total of 45 million people were affected, and more than 800 billion economic losses were caused. It can be seen that major geological disasters have a large influence range, a large number of affected population, and serious economic losses (including buildings, roads, cash crops, etc.), which

can have a serious impact on people's lives and even cause devastating consequences. In recent years, the occurrence of major earthquake disasters is more frequent, resulting in serious losses and difficult rescue, which is the focus and difficulty of emergency management research. Once a major earthquake disaster occurs, its loss and impact are immeasurable. It is necessary to strengthen the work of emergency rescue and response to major earthquake disasters. In the process of emergency rescue of major earthquake disasters, the core is the rescue of disaster victims, and the main participants are rescue teams. The effective deployment of rescue teams and the cooperation between rescue teams are important to guarantee to control the further deterioration of disaster situation and reduce casualties in disaster areas [1, 2].



Foreign experts and scholars have conducted a lot of research on emergency disaster rescue, and many research results are worth learning. Based on literature research and expert research, Baroni first selected the relevant important indicators to evaluate the priority and obtained the weight of these indicators through an analytic hierarchy process (AHP). Then, the spatial multicriteria decision analysis method was used for modeling in GIS to analyze the priority of disaster areas. Finally, the urban search and rescue operations were allocated based on this. But, his research is not applicable to earthquake disaster relief. Nakanishi et al. discussed the influence of the environment on each individual in pedestrian evacuation and proposed a social force evacuation model. In order to better describe the force of individual pedestrian evacuation, the model established three indicators including self-driving force, surrounding pedestrian influence, and surrounding obstacle influence. However, their conclusion is not supported by specific experimental data [3]. Gordon uses a mathematical model to establish an emergency plan and establishes an emergency decision support system of automatic response, which realizes the automatic decision making of the system through the interdependence and restriction relationship among the response tasks in the emergency plan. However, his contingency plan is not comprehensive enough.

Based on the theory of complex adaptive system, this paper establishes the earthquake disaster rescue model. By analyzing the task allocation mechanism and team deployment model, this paper studies the optimization method of emergency deployment decision making of a rescue team, which can effectively save the wounded and minimize the loss caused by the disaster, so as to realize the improvement of the emergency rescue effect in a real sense and provide a reliable basis for disaster relief and disaster reduction work.

## 2. Theoretical Basis of the Complex Adaptive System and Establishment of the Disaster Rescue Model

### 2.1. Complex Adaptive System Theory

*2.1.1. Theoretical Basis of the Complex Adaptive System.* The complex adaptive system is composed of adaptive agents, who constantly learn and accumulate experience in the process of interaction and accordingly change their own structure and behavior mode [4]. It is this kind of adaptive behavior that makes the subject and the environment constantly change, so that the system constantly evolves and becomes more complex. The theory of complex adaptive system includes two levels: a macrolevel and microlevel. The microlevel refers to that the adaptive subject constantly adjusts the behavior criterion according to the effect of behavior in the process of interaction with the environment to realize adaptive survival; the macrolevel refers to the system composed of adaptive subjects, which realizes adaptation, emergence, and differentiation in the process of interaction between the main body and the external environment iteration.

Professor Holland proposed that the complex adaptive system includes four characteristics, aggregation, flow, nonlinearity, and diversity, and three mechanisms, identification, internal model, and building blocks [5]. Generally speaking, all the systems that meet the abovementioned seven basic points are complex adaptive systems and can be applied to solve related problems.

*2.1.2. Complex Adaptive Subject.* The complex adaptive system is composed of multiple individuals who actively and intelligently interact to realize the evolution and iteration of the system. Adaptation refers to the active and repeated interaction between the subject and the environment. The so-called subject refers to the active individuals in the system. The subjects aggregate into larger subjects and then form the system. Therefore, the subjects at different levels are also systems in different levels. When the adaptive agent plays an adaptive role, it follows the common model system rules [6, 7]:

#### (1) Executive system model

The executive system model is the “detector input + IF/THEN rule set + effector output” model. The detector filters the information in the surrounding environment, receives the useful stimulus, and transmits it to the subject, which reflects the agent’s ability to collect information. The IF/THEN rule set is a set of rules that define the subject. After receiving the information from the detector, the agent uses the rule set to process, continuously activate other rules, or directly activate the effector, reflecting the agent’s ability to analyze and process information. After the effector is activated, the subject takes corresponding actions to reflect the subject’s ability to cope with environmental changes.

#### (2) Credit assignment

When dealing with and applying rules, adaptive agents will comprehensively consider the environment and other agents and rank the roles of various rules. The assignment of this role is the trust degree. The competitiveness of a rule, that is, the trust degree, mainly depends on the past usefulness of the rule. In the process of adapting to the environment, the usefulness of various rules is changing and the trust degree is also changing. Therefore, the process of modifying rule trust strength based on experience and learning is credit assignment.

#### (3) Rule generation

The environment of the complex adaptive system is complex and changeable, and the adaptability of the main body is also changing. Any IF/THEN rule set is accumulated and evolved in adaptive activities. The existing rules with high success rate generate new rules by copying and reorganizing, thus enriching the rule set and increasing the fitness of the subject. The continuous emergence of new rules is an important source for the subject to adapt to the

environment and is also the main driving force of system evolution iteration.

**2.1.3. Complex Adaptive System Model.** The evolution process of the complex adaptive system is realized by the stimulation response model in the microlevel and the echo model in the macrolevel. The stimulus-response model is for adaptive agents, while the echo model is for complex adaptive systems. The model application of CAS has the characteristics of initiative, dynamic operability, and hierarchy. It studies the evolution process of CAS in a way that is more in line with the facts and combines qualitative and quantitative changes [8].

(1) *Stimulus-Response Model.* The stimulus-response model describes how the subject adapts, learns, and accumulates experience. The adaptive agent can respond to the stimulation of the surrounding environment and other agents by establishing the executive system model, credit allocation, and rule generation, thus evolving towards the highest point of adaptability in many directions.

(2) *Echo Model.* The echo model is based on the stimulus-response model, which simulates, describes, and studies the behavior of the whole complex adaptive system from a macroscopic perspective. In the echo model, in addition to defining topics, sites and resources are defined. The complex adaptive system is composed of several sites, which are connected with each other; each site has resource and environmental conditions to accommodate several main activities; the subject site carries out resource and information exchange. The basic model of the echo model can describe the resource exchange activities of adaptive agents at the same level and basically reflect the functional relationship and behavior mechanism between adaptive agents.

In the basic model, the agent is composed of a resource database, attack identifier, and defense identifier. The resource library processes and stores the acquired resources; it is the attack mark that actively establishes contact with other subjects to explore whether there are needed resources; the defense identification is used to accept the contact of other subjects, which is used to respond when receiving the contact of other subjects [9]. The basic model of the complex adaptive system is shown in Figure 1.

## 2.2. Overview of Earthquake Disaster Rescue Process

**2.2.1. Disaster Emergency Rescue Process.** Disaster emergency rescue refers to a series of means and countermeasures adopted by the government and other public organizations in the process of prevention, response, response, and recovery in the process of sudden and destructive emergency disasters. The purpose is to ensure the life safety of the people in the disaster area as much as possible, so as to minimize the loss caused by the disaster [10, 11]. The three core links of disaster emergency rescue process are “preparation before disaster,” “emergency in disaster,” and “recovery after disaster.”

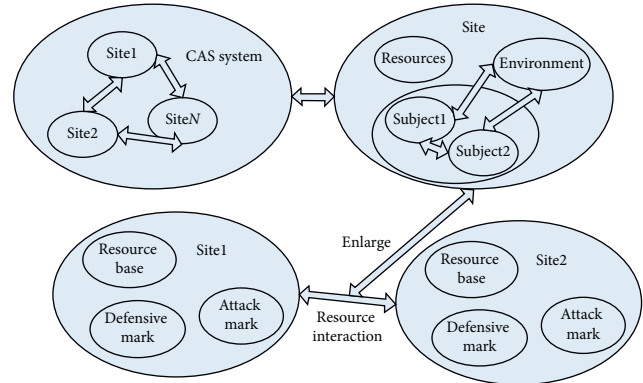


FIGURE 1: Basic model of the complex adaptive system.

### (1) Predisaster preparation stage

The purpose of predisaster preparation is to ensure the necessary conditions and capabilities for post-disaster emergency rescue work. The main task is to take some organizational preparations and emergency support in advance for possible disaster events, so as to ensure the response and disposal in the first time after the occurrence of disasters and prevent the deterioration of disaster emergencies. The main work includes the distribution of emergency resources and the distribution of emergency materials emergency plan design, etc.

### (2) Emergency stage in disaster

The emergency stage in disaster refers to the rescue measures taken after the occurrence of major geological disasters in order to prevent further expansion or deterioration of disaster events and minimize the loss or damage caused by disasters. The main research contents of the emergency stage include the demand forecast of emergency resources, the allocation of emergency resources, personnel search, and rescue and evacuation.

### (3) Postdisaster recovery stage

Postdisaster recovery refers to the measures or actions taken to restore production, work, life, and ecological environment to the normal state as soon as possible after the impact of disaster events is controlled. At present, the research on the recovery stage mainly focuses on the assessment of emergency response capacity.

### 2.2.2. Task Allocation of Disaster Search and Rescue.

Most of the research studies on task allocation are related to work allocation and cooperation in a multirobot system [12]. In fact, the research of task allocation can also be applied to solve some similar problems in many other fields, such as UAV cooperation, RoboCup, assignment problem, and emergency rescue cooperation. The main factors of disaster search and rescue task allocation and classification are the complexity of the task and the function of the main body. The tasks to be completed are divided into simple tasks and

complex tasks according to their complexity. Simple tasks represent tasks that only need one agent to complete, and complex tasks represent tasks that need to be completed by multiple agents. According to the function, the subject of task execution is divided into single task subject and multitask subject. Single task subject means that the subject can execute at most one task at the same time, and multitask subject means that the subject can perform multiple tasks at the same time [13, 14].

(1) Single task subject- simple task

Task allocation in the context of single task subject and simple task is actually a classic assignment problem. Assuming that there are  $m$  single task agents  $a_i$  and  $n$  single agent tasks  $t_j$ , each agent can complete any task with corresponding cost  $c_{ij}$  and completion quality  $q_{ij}$ . The goal of the assignment model can be to minimize the total cost after allocation or to achieve the highest total completion quality or to maximize the allocation benefit (a comprehensive consideration of the total cost and completion quality). Suppose  $m = n$ ; then, this is a balanced assignment problem.

$$\begin{aligned} \text{Maximize } Z &= \sum_{i=1}^m \sum_{j=1}^n x_{ij} u_{ij}, \\ \text{s.t. } \sum_{i=1}^m x_{ij} &= 1, \quad 1 \leq j \leq n, \\ \sum_{j=1}^n x_{ij} &= 1, \quad 1 \leq i \leq m. \end{aligned} \quad (1)$$

(2) Single task subject- complex task

Some of these problems can be transformed into set partitioning problems. Consider a nonempty set  $Y$ ; the set  $Y$  is divided into several nonempty subsets, these nonempty subsets have no intersection, and the union of the elements of these nonempty subsets is exactly equal to  $Y$ . If  $x$  is used to represent the set of nonempty subsets, there exists a utility function:

$$u: X \longrightarrow R^+. \quad (2)$$

Then, we need to find the set partition method  $X$  to maximize the utility. SPP is a NP hard problem, but on the other hand, many heuristic algorithms have been proposed to solve the problem.

(3) Multitask subject- simple task

This kind of problem is not very common in reality because it requires a single agent to perform multiple tasks at the same time, which only appears in a very few cases. However, the method to solve this kind of problem is similar to the single task subject complex task problem mentioned above. Some algorithms such as spp can be applied to these two models.

(4) Multitask subject- complex task

The task assignment problem can be transformed into the collection coverage problem. Set  $Y$  as a nonempty set. The elements in set  $Y$  can form several nonempty subsets, and these nonempty subsets can intersect. If  $X$  is used to represent the set of these nonempty subsets, then there is a cost function:

$$c: X \longrightarrow R^+. \quad (3)$$

Then, we need to find a subset of  $X$  that minimizes the cost, and the union of the elements in these subsets is exactly equal to  $Y$ . The SCP problem is also NP hard, but many scholars have proposed the algorithm to obtain the approximate solution.

*2.3. Earthquake Disaster Rescue Deployment Model.* After the occurrence of large-scale earthquake disaster, there are  $j$  disaster areas with scattered geographical locations and different disaster situations and  $i$  rescue points participate in the rescue. There are rescue teams with different numbers and abilities in each rescue point. For any disaster site, according to the disaster information we collected, we can know the type of rescue team that the disaster site needs the most. Therefore, the optimal deployment of rescue teams can be transformed into the problem of finding the most similar rescue teams. By calculating the similarity between rescue teams and selecting the rescue teams with high similarity for deployment, the emergency rescue effect is the best [15].

According to the actual situation of emergency rescue after a large-scale earthquake disaster, the following assumptions are given: each rescue team may include four types of personnel, namely, armed police officers and soldiers, doctors, nurses, and volunteers; the number of rescue teams required by the disaster site is highly related to the population size and the number of survivors; the road condition and corresponding geographic location information are obtained through GIS.

Once a large-scale earthquake disaster occurs, it often causes damage to the road network in the disaster area. Combined with the internationally accepted definition of the 72 hour golden rescue period after the disaster, in order to shorten the rescue time, the helicopter is also used as an emergency transport tool [16]. The parameter  $t_{ikj}$  represents the transportation time from the  $k$ -type transport vehicle in the rescue point  $i$  to the disaster site  $j$  under the condition of road network damage. Considering the characteristics of the transport vehicle, it is discussed in the following situations:

- (1) The transport tool  $k \in H.$ , in the rescue point, is transported by a helicopter. At this time, the transportation time is not affected by the road conditions; therefore,  $t_{ikj} = t_{ikj}^n$ .
- (2) The means of transport  $k \in E.$ , in the rescue point, that is, road transportation, is adopted. At this time, the transportation time will be affected by the road conditions. According to the road damage degree  $\lambda_{ij}$ , it can be divided into the following situations:

- (1) When  $\lambda_{ij}=0$ , the road is not damaged and the transportation time is not affected,  $t_{ikj} = t_{ikj}^n$ .
- (2) When  $0 < \lambda_{ij} < \alpha$ , it means that the road has been damaged but can pass and the transportation time is affected. At this time,  $t_{ikj} = t_{ikj}^n (1 + \lambda_{ij})$ .
- (3) When  $\alpha \leq \lambda_{ij} \leq 1$ , it means that the road is damaged and cannot be passable, but can be repaired. At this time, the transportation time is  $t_{ikj} = t_{ikj}^n (1 + \lambda_{ij}) + t_{ij}^r$ .
- (4) When  $\lambda_{ij} = 1$ , it means that the road is damaged and cannot be repaired in a short time, so the route is not feasible and the transportation time is  $t_{ikj} = +\infty$ .

To sum up, the parameter  $t_{ikj}$  can be expressed as follows:

$$t_{ikj} = \begin{cases} t_{ikj}^n, & k \in H. \text{ or } k \in E. \text{ and } \lambda_{ij} = 0, \\ t_{ikj}^n (1 + \lambda_{ij}), & k \in E. \text{ and } 0 < \lambda_{ij} < \alpha, \\ t_{ikj}^n (1 + \lambda_{ij}) + t_{ij}^r, & k \in E. \text{ and } \alpha \leq \lambda_{ij} < 1, \\ +\infty, & k \in E. \text{ and } \lambda_{ij} = 1. \end{cases} \quad (4)$$

Based on the abovementioned analysis, this paper establishes a rescue team deployment model to solve the contradiction between the disaster severity and the complexity of the rescue team in the process of large-scale earthquake disaster rescue team deployment, effectively depict the difference between the rescue team and disaster site, and realize the personalized deployment of the rescue team.

After the occurrence of a large-scale earthquake disaster, the earlier the rescue team arrives in the disaster area, the more we will be able to reduce casualties and property losses. In addition, the rescue team is different from the emergency

materials, and the same rescue team deployed to different disaster areas has different effects [17, 18]. Therefore, the reasonable deployment of a large-scale earthquake disaster rescue team must consider both timely and efficient objectives, that is, to deploy the most suitable rescue team to the disaster site in the shortest time.

- (1) Shortest time target of emergency rescue

The time the rescue team arrives in the disaster area will directly affect the rescue effect. Therefore, the shortest emergency rescue time target is an important objective function in the deployment model of a large-scale geological disaster rescue team. This paper uses the time when all rescue teams arrive at the disaster area to measure [19]. Since the rescue vehicles in each rescue point can be parallel, the maximum rescue transportation time is taken as the minimum.

- (2) Goal of maximizing the effectiveness of rescue team deployment

The difference between the rescue team and the disaster site leads to the different effects of the same rescue team deployed to different disaster areas. Therefore, in the process of emergency deployment, we should strive to make the best use of the talents of the people and make the best use of the personnel. In this paper, the complex adaptive system is used to realize the reasonable deployment of rescue teams, that is, to maximize the effectiveness of the deployment of rescue teams.

On this basis, the deployment model of the rescue team based on the complex adaptive system is constructed as follows:

$$\text{Min } Z_1 = \max_{j \in J} \left\{ \max_{i \in I} \left\{ \max_{k \in K_i} t_{ikj} \left( 2 \left\lceil \frac{(\sum_{p \in P_i} y_{ikpj} * n_{ip})}{q_k} \right\rceil - 1 \right) \right\} \right\}, \quad (5)$$

$$\text{Max } Z_2 = \sum_{j \in J} \sum_{i \in I} \sum_{k \in K_i} \sum_{p \in P_i} e_{ipj} y_{ikpj}, \quad (6)$$

$$\text{s.t. } \sum_{j \in J} \sum_{k \in K_i} \sum_{p \in P_i} y_{ikpj} \leq s_i, \quad i \in I, \quad (7)$$

$$\sum_{i \in I} \sum_{k \in K_i} \sum_{p \in P_i} y_{ikpj} \leq d_j, \quad j \in J, \quad (8)$$

$$\sum_{k \in K_i} \sum_{j \in J} y_{ikpj} = 1, \quad i \in I, p \in P_i, \quad (9)$$

$$\sum_{i \in I} \sum_{k \in K_i} \sum_{p \in P_i} \sum_{j \in J} c_p y_{ikpj} + \sum_{i \in I} \sum_{k \in K_i} \sum_{j \in J} 2c_{ikj} \left\lceil \frac{(\sum_{p \in P_i} y_{ikpj}) * n_{ip}}{q_k} \right\rceil \leq C_0, \quad (10)$$

$$\sum_{i \in I} \sum_{k \in K_i} \sum_{p \in P_i} \sum_{j \in J} y_{ikpj} > \varphi, \quad (11)$$

$$y_{ikpj} \in \{0, 1\}, \quad i \in I, j \in J, p \in P_i, \quad (12)$$

$$t_{ikj} = \begin{cases} t_{ikj}^n, & k \in H. \text{ or } k \in E. \text{ and } \lambda_{ij} = 0, \\ t_{ikj}^n(1 + l\lambda_{ij}), & k \in E. \text{ and } 0 < \lambda_{ij} < \alpha, \\ t_{ikj}^n(1 + l\lambda_{ij}) + t_{ij}^r, & k \in E. \text{ and } \alpha \leq \lambda_{ij} < 1, \\ +\infty, & k \in E. \text{ and } \lambda_{ij} = 1. \end{cases} \quad (13)$$

Formulas (7) to (13) are the constraints of the model, where formula (7) indicates that the deployment quantity of rescue teams does not exceed the available quantity; formula (8) indicates that the deployment quantity of rescue teams does not exceed the demand of the disaster stricken area; formula (9) indicates that each rescue team can only be deployed to one disaster site no matter what transportation mode is used; and formula (10) represents the constraint of available rescue cost at the initial stage after the disaster, which reflects the possible financial constraints in the early postdisaster period, and the emergency manager can adjust the parameter  $C_0$  according to the actual situation; formula (11) indicates that at least 0 rescue teams should be deployed to participate in the rescue work, and  $\varphi$  is a constant greater than 1; equation (12) is a decision variable, indicating whether to deploy rescue teams; and equation (13) represents the transportation time from the rescue point to the disaster site when the road network is damaged.

### 3. Rescue Model Simulation Experiment Based on the Complex Adaptive System Theory

**3.1. Comparison Objects.** This paper establishes an earthquake disaster rescue model based on the theory of complex adaptive systems. In order to verify the performance and effectiveness of this model, it is first compared with other rescue models and then compared with the rescue model based on the *F*-Max-Sum algorithm. *F*-Max-Sum is also a distributed algorithm, which is improved on the basis of Max-Sum algorithm. It is suitable for solving similar dynamic task allocation problems and has excellent performance [20, 21]. At present, the Max-Sum algorithm and the *F*-Max-Sum algorithm are used for task allocation in disaster environments, as well as for the allocation of spectrum resources and cloud resources. The *F*-Max-Sum algorithm first converts the problem to be solved into the expression form of a factor graph and iterates the value of information transfer between variable nodes and function nodes in the factor graph until it converges or iterates a certain number of times.

**3.2. Experimental Parameter Setting.** This paper uses *R* and Netlogo software to calculate and simulate. It is assumed that the disaster area is a circular area with a radius of 150 m, and

a certain number of wounded are scattered in several burial sites. The simulation step size ticks are in minutes. When there are no survivors on the scene, the simulation ends. The longest simulation cycle is 72 hours.

Combined with the statistical results of earthquake casualties in recent years, assuming the specific distribution of the injury degree of landslide disaster under different scenarios, as shown in Table 1, in the “extremely serious” scenario, the proportion of wounded death is 35%, the proportion of serious injury is 30%, the proportion of minor injury is 20%, and the proportion of no injury is 15%.

The input variables include the number of wounded in each mask point  $n_s-U(1,3)$ ; the injury degree  $v_{si}$  of each wounded person is given the initial number according to Table 1; and the burying condition of each buried point is  $v_{\text{rubble}}-N(120,30)$ ; the results show that the search speed is 3 m/min; the radius shown is  $s_r = 3$  m; and the rescue speed is  $S_{\text{rubble}} = 0.5$ ; when there is a buried point in the search radius, the simulation rescue team can find it with a certain probability, assuming the probability of prob = 20%; the cooperation range  $s_c$  is 40 m.

### 4. Simulation Results and Performance Comparison of Different Rescue Models

**4.1. Performance Comparison of Different Rescue Models.** As shown in Table 2 and Figure 2, the algorithm in the model used in this paper converges after 40 hours of running the simulation program, while the *f*-max-sum model converges after 50 hours. The simulation results show that the convergence speed of the model algorithm in this paper is faster than that of the other two model algorithms. The reason is that, firstly, the model algorithm in this paper does not blindly process all the perceived information, but only selects the information that has an impact on the assistance task. Secondly, it uses the screened information to predict the relevant state and formulate the cooperation strategy, which can quickly find the convergence path after interference, so as to improve the rescue efficiency.

As shown in Table 3 and Figure 3, it can be seen that, after 72 hours, the rescue number of the model in this paper reached 138, higher than that of the other two models. The number of rescuers in this model is higher than that of the other two models because the cooperation information obtained is more comprehensive, which improves the accuracy of the prediction results, and uses action evaluation

TABLE 1: Distribution of the injury degree under different disaster scenarios.

| Degree of injury- disaster scenario | Extremely serious (%) | Serious (%) | Ordinary (%) |
|-------------------------------------|-----------------------|-------------|--------------|
| Death                               | 35                    | 25          | 15           |
| Serious injury                      | 30                    | 20          | 20           |
| Minor wound                         | 20                    | 35          | 25           |
| No injuries                         | 15                    | 20          | 40           |

TABLE 2: Comparison of convergence of different model algorithms.

| Path length        | 10 H | 20 H | 30 H | 40 H | 50 H | 72 H |
|--------------------|------|------|------|------|------|------|
| This paper         | 34   | 23   | 14   | 9    | 9    | 9    |
| Other model        | 38   | 30   | 23   | 16   | 12   | 9    |
| $F$ -max-sum model | 37   | 25   | 19   | 14   | 9    | 9    |

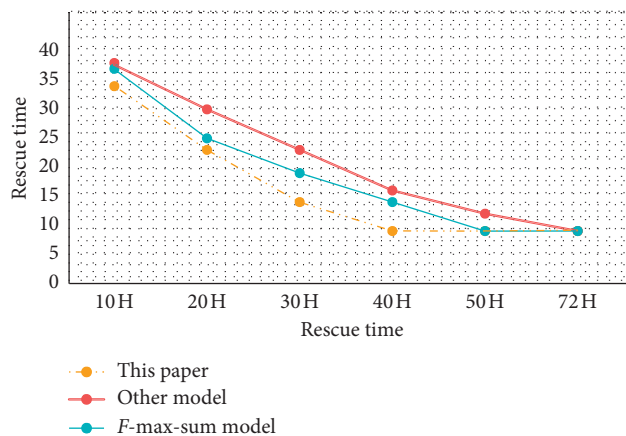


FIGURE 2: Comparison of convergence of different model algorithms.

and trigger class to realize the dynamic adjustment of the cooperation strategy to help the task to be completed smoothly.  $F$ -max-sum model algorithm reduces the efficiency of cooperation due to the high complexity of information processing algorithm, and it is difficult to get the optimal cooperation strategy, which leads to low rescue efficiency.

As shown in Table 4 and Figure 4, the overall scores of the three models in the simulation experiment within 72 hours are shown. Among them, the score of the model proposed in this paper is the highest among the three models after 72 hours, reaching 98. The scores of the other two models are 80 and 84, respectively. When the simulation program runs for about 40 hours, the scores of the three models are the closest. The results in Table 4 and Figure 4 further show that the performance of the proposed model is better than that of other models.

*4.2. Comparison of Simulation Results of Different Rescue Models.* As shown in Table 5 and Figure 5, the simulation results of the rescue model in this paper are compared with those of the other two models, and some indexes of the former are better than those of the latter two. The results

show that compared with other models, the performance of this model is not inferior, and some indicators are better than the other two models. One of the biggest advantages of this model is the low complexity of the calculation process. In the iterative process of  $f$ -max-sum algorithm, function nodes need to traverse the various combinations of surrounding variable nodes, which will consume a lot of time, especially when the cooperation set is large, that is, there are more rescue teams to be deployed around the rescue task. In addition,  $f$ -max-sum was originally designed for the cooperation between devices with computing functions, such as robots and sensors. It has adaptability problems when it is directly applied to the cooperation between rescue teams on the disaster scene (Tables 6 and 7).

As shown in Figures 6 and 7, the average situation of the relative production probability and the proportion of the wounded rescued in the earthquake disaster rescue based on the complex adaptive system within about 2000 simulation steps in three different scenarios after 100 simulation times are shown. It can be seen that, in the initial stage of rescue, the relative survival probability decreases rapidly. With the development of rescue work, the decline speed of relative survival probability slows down after about 300 minutes. In the early stage of rescue, the average survival probability

TABLE 3: Comparison of rescue numbers of different models.

| Number of rescuers      | 10 H | 20 H | 30 H | 40 H | 50 H | 72 H |
|-------------------------|------|------|------|------|------|------|
| This paper              | 28   | 40   | 77   | 101  | 130  | 138  |
| Other model             | 18   | 43   | 71   | 106  | 120  | 123  |
| <i>F</i> -max-sum model | 19   | 37   | 60   | 108  | 124  | 126  |

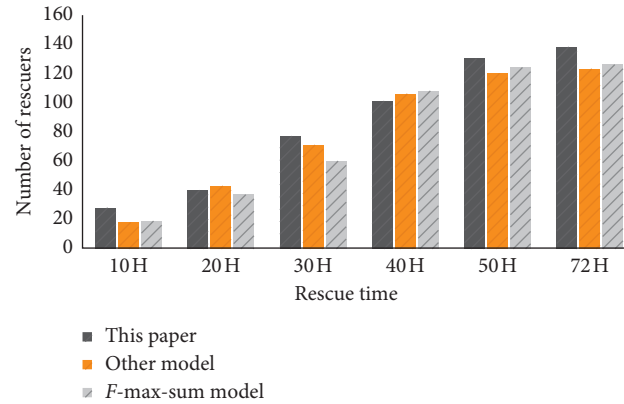


FIGURE 3: Comparison of rescue numbers of different models.

TABLE 4: Comparison of overall scores of three models.

| Overall score           | 10 H | 20 H | 30 H | 40 H | 50 H | 72 H |
|-------------------------|------|------|------|------|------|------|
| This paper              | 24   | 33   | 56   | 72   | 91   | 98   |
| Other model             | 19   | 37   | 49   | 76   | 80   | 80   |
| <i>F</i> -max-sum model | 16   | 38   | 51   | 75   | 81   | 84   |

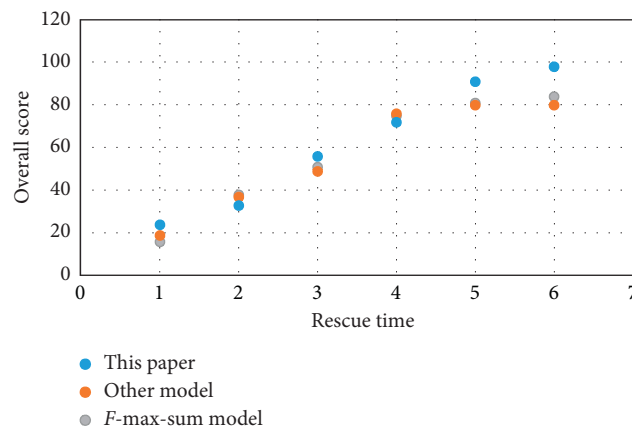


FIGURE 4: Comparison of overall scores of three models.

TABLE 5: Comparison of simulation results of three rescue schemes.

|                   |                         | Relative survival probability (%) | Proportion of wounded rescued (%) | Average rescue time (min) |
|-------------------|-------------------------|-----------------------------------|-----------------------------------|---------------------------|
| Extremely serious | This model              | 58.92                             | 66.31                             | 523.1                     |
|                   | <i>F</i> -max-sum model | 54.38                             | 62.48                             | 541.6                     |
|                   | Other model             | 48.74                             | 59.18                             | 584.2                     |
| Serious           | This model              | 67.85                             | 76.45                             | 596.1                     |
|                   | <i>F</i> -max-sum model | 65.17                             | 74.21                             | 610.7                     |
|                   | Other model             | 61.94                             | 70.94                             | 645.6                     |
| Ordinary          | This model              | 77.46                             | 83.06                             | 631.3                     |
|                   | <i>F</i> -max-sum model | 74.37                             | 83.53                             | 659.7                     |
|                   | Other model             | 71.25                             | 82.17                             | 684.6                     |

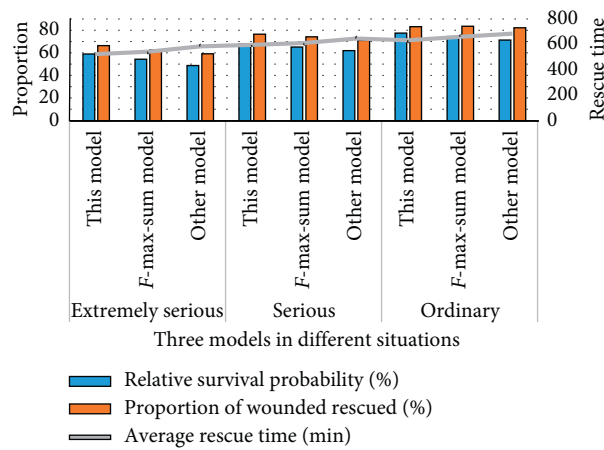


FIGURE 5: Comparison of simulation results of three rescue schemes.

TABLE 6: Changes of survival rate in different situations.

|                   | 1 | 501  | 1001 | 1501 | 2001 |
|-------------------|---|------|------|------|------|
| Extremely serious | 1 | 0.57 | 0.56 | 0.55 | 0.54 |
| Serious           | 1 | 0.7  | 0.7  | 0.69 | 0.68 |
| Ordinary          | 1 | 0.83 | 0.8  | 0.78 | 0.76 |

TABLE 7: Proportion of wounded rescued in different situations.

|                   | 1    | 501  | 1001 | 1501 | 2001 |
|-------------------|------|------|------|------|------|
| Extremely serious | 0.42 | 0.67 | 0.83 | 0.98 | 1    |
| Serious           | 0.42 | 0.66 | 0.82 | 0.9  | 0.93 |
| Ordinary          | 0.42 | 0.65 | 0.81 | 0.83 | 0.85 |

decreases faster, mainly because of the existence of seriously injured trapped people. However, the survival probability of such wounded people drops very fast, about 300 minutes. After that, the seriously injured died or were rescued and only slightly injured or not injured persons were left on the whole scene, and the average survival probability of the wounded decreased slowly.

*4.3. Frequency of Earthquakes in Recent Years.* We have made statistics on the frequency of earthquakes with magnitude 5 and above in recent years and also the property and life damage caused by them, as shown in Table 8.

It can be seen from Figure 8 that, in recent years, the global threat of earthquakes has been more serious. On average, there are about 5 earthquakes of magnitude 7 or



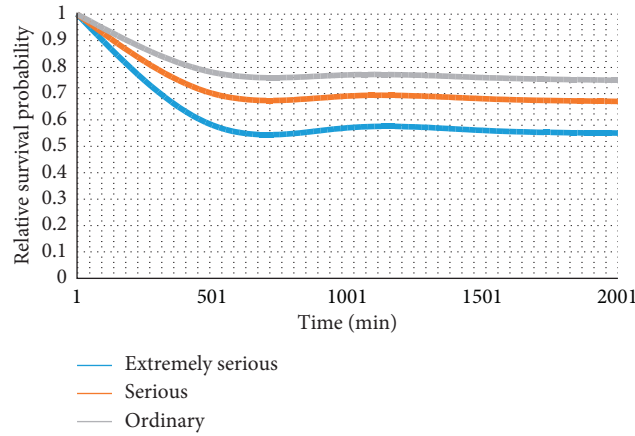


FIGURE 6: Changes of survival rate in different situations.

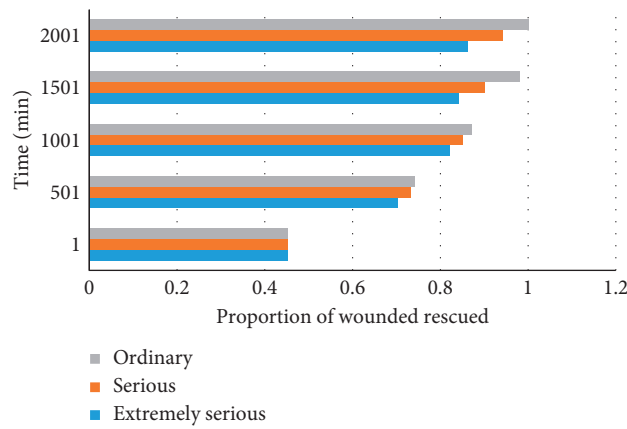


FIGURE 7: Proportion of wounded rescued in different situations.

TABLE 8: Number of earthquakes worldwide.

|      | Level 3 | Level 4 | Level 5 | Level 6 | Level 7 |
|------|---------|---------|---------|---------|---------|
| 2015 | 165     | 53      | 18      | 16      | 7       |
| 2016 | 177     | 42      | 24      | 14      | 5       |
| 2017 | 162     | 47      | 26      | 17      | 3       |
| 2018 | 159     | 41      | 22      | 15      | 9       |
| 2019 | 163     | 39      | 15      | 14      | 6       |

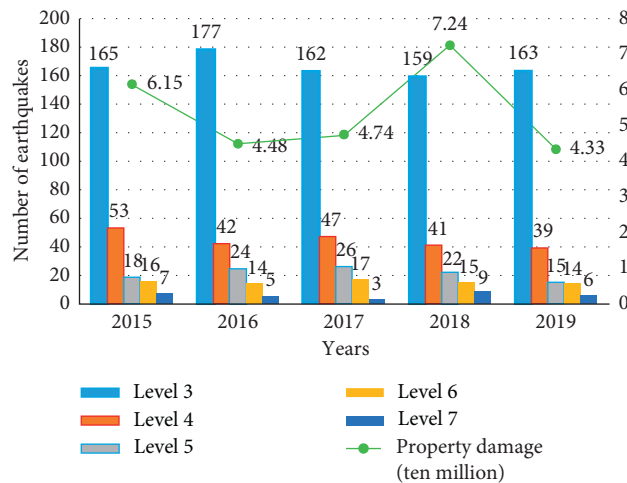


FIGURE 8: Number of earthquakes and losses.

above each year. The remaining small earthquakes are even harder to count. The property losses caused by earthquakes are as high as tens of millions each year. This also shows that the earthquake relief system proposed in this article is particularly necessary.

## 5. Conclusions

The frequent occurrence of major earthquake disasters is an urgent problem faced by all countries in the world. The traditional emergency management and emergency response decision making focus on the dispatching and distribution of relief materials, ignoring the role of human beings. In fact, in the process of disaster relief, the rescue team is the main body of participation. Under the background of frequent occurrence and increasingly serious impact of major earthquake disasters, it is not only of important theoretical value to carry out the research on search and rescue model to deal with such disasters but also has a strong practical significance for guiding the emergency rescue work after disasters and improving the effectiveness of the rescue decision-making scheme.

The main research content of this paper is an earthquake disaster rescue model based on the theory of complex adaptive systems. Aiming at the contradiction between the severity of large-scale earthquake disasters and the complexity of rescue teams, an earthquake disaster rescue model based on complex adaptive systems is established. The model considers the two goals of timeliness and efficiency, deploys the most suitable rescue team to the disaster site in the shortest time, and better describes the needs of large-scale geological disaster emergency rescue decision making. In the context of large-scale earthquake disasters, the research content of this paper solves many difficulties faced in the deployment of rescue teams and improves the efficiency of emergency rescue work. It provides an effective solution tool for postdisaster emergency rescue decision making and improves the emergency response capability for large-scale geological disasters.

Although this paper has carried on the beneficial exploratory research on the optimization deployment and decision making of a large-scale earthquake disaster rescue team and has achieved certain innovative results, there are still many work that need further study, mainly including the following: in the future, we will study the demand quantity of the rescue team in a disaster area by the tensor decomposition and filling method; pay attention to the data collection of rescue cases, according to the model proposed by relevant scholars, and use limited data to correct the model parameters, so as to improve the applicability of the model to the actual situation.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] W. Heping, H. Weihua, T. Ming et al., "A preliminary analysis of the earthquake disaster of buildings during two destructive earthquakes in Xinjiang," *Earthquake Research in China*, vol. 32, no. 3, pp. 117–129, 2018.
- [2] H. T. Wong and S. Li, "Healthcare services demand in post-disaster settings: the 2014 earthquake in Ludian county, Yunnan province, China," *International Journal of Disaster Risk Science*, vol. 7, no. 4, pp. 445–449, 2016.
- [3] T. Nakanishi, M. Takeyama, and A. Noya, "Occupational health and security protection of workers on nuclear and radiation emergency rescue," *Chinese Journal of Radiological Health*, vol. 112, no. D6, pp. 541–553, 2015.
- [4] C. Coetzee, D. Van Niekerk, and E. Raju, "Disaster resilience and complex adaptive systems theory- finding common grounds for risk reduction," *Disaster Prevention and Management*, vol. 25, no. 2, pp. 196–211, 2016.
- [5] Z. Yang, M. Yin, J. Xu, and W. Lin, "Spatial evolution model of tourist destinations based on complex adaptive system theory: a case study of Southern Anhui, China," *Journal of Geographical Sciences*, vol. 29, no. 8, pp. 1411–1434, 2019.
- [6] A. J. Burns, C. Posey, J. F. Courtney et al., "Organizational information security as a complex adaptive system: insights from three agent-based models," *Information Systems Frontiers*, vol. 19, no. 3, pp. 1–16, 2017.
- [7] K. Zhang and A. H. J. Schmidt, "Thinking of data protection law's subject matter as a complex adaptive system: a heuristic display," *Computer Law & Security Review*, vol. 31, no. 2, pp. 201–220, 2015.
- [8] P. Mcevoy, M. Brady, and R. Munck, "Capacity development through international projects: a complex adaptive systems perspective," *International Journal of Managing Projects in Business*, vol. 9, no. 3, pp. 528–545, 2016.
- [9] M. Janssen, H. van der Voort, and A. F. van Veenstra, "Failure of large transformation projects from the viewpoint of complex adaptive systems: management principles for dealing with project dynamics," *Information Systems Frontiers*, vol. 17, no. 1, pp. 15–29, 2015.
- [10] X. Lin, L. Zhai, M. Zhang, Y. Wang, and J. Li, "Ergonomic evaluation of protective clothing for earthquake disaster search and rescue team members," *International Journal of Clothing Science and Technology*, vol. 28, no. 6, pp. 820–829, 2016.
- [11] Y. J. Zheng, Q. Z. Chen, H. F. Ling et al., "Rescue wings: mobile computing and active services support for disaster rescue," *IEEE Transactions on Services Computing*, vol. 9, no. 4, p. 1, 2016.
- [12] Y. Jiang, "A survey of task allocation and Load balancing in distributed systems," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 2, pp. 585–599, 2016.
- [13] S. S. Jha and S. B. Nair, "TANSA: task allocation using nomadic soft agents for multirobot systems," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 4, pp. 308–318, 2018.
- [14] S. Mayya, S. Wilson, and M. Egerstedt, "Closed-loop task allocation in robot swarms using inter-robot encounters," *Swarm Intelligence*, vol. 13, no. 2, pp. 115–143, 2019.
- [15] X. Hu, Z. Liu, Y. Yao et al., "Crowdsourcing model research for the identification of post-earthquake rescue Objects," *Journal of Earthquake Engineering*, vol. 23, no. 3–5, pp. 863–881, 2017.

- [16] T. Bloch, R. Sacks, and O. Rabinovitch, "Interior models of earthquake damaged buildings for search and rescue," *Advanced Engineering Informatics*, vol. 30, no. 1, pp. 65–76, 2016.
- [17] J. Xu, Z. Wang, M. Zhang et al., "A new model for a 72-h post-earthquake emergency logistics location-routing problem under a random fuzzy environment," *Transportation Letters the International Journal of Transportation Research*, vol. 8, no. 5, pp. 1–16, 2016.
- [18] B. Wei, G. Nie, G. Su et al., "Risk assessment of people trapped in earthquake based on km grid: a case study of the 2014 Ludian earthquake, China," *Geomatics, Natural Hazards and Risk*, vol. 8, no. 2, pp. 1–17, 2017.
- [19] Y. Nagasaka, A. Nozu, and A. Wakai, "Strong ground motion simulation for the 2005 Central Chiba prefecture earthquake with pseudo point-source model," *Journal of Japan Association for Earthquake Engineering*, vol. 16, no. 3, pp. 3\_17–3\_29, 2016.
- [20] Z. Chen, Y. Deng, T. Wu, and Z. He, "A class of iterative refined Max-sum algorithms via non-consecutive value propagation strategies," *Autonomous Agents and Multi-Agent Systems*, vol. 32, no. 6, pp. 822–860, 2018.
- [21] R. Hadfi and T. Ito, "Low-complexity exploration in utility hypergraphs," *Journal of Information Processing*, vol. 23, no. 2, pp. 176–184, 2015.

## Research Article

# Construction of a Complex System Based on Big Data for the Intelligent Service System of Youth Physical Health

Xinwen Li,<sup>1</sup> Chao Song,<sup>2</sup> Christine A. Rochester,<sup>3</sup> and Chaobing Yan <sup>4,5</sup>

<sup>1</sup>Department of Physical Education, University of Electronic Science and Technology of China, Chengdu 610054, Sichuan, China

<sup>2</sup>Department of Weapons and Control, Army Academy of Armored Forces, Beijing 100072, China

<sup>3</sup>Department of Exercise Science, Physical Education and Recreation of Colorado State University, Pueblo, CO 81001, USA

<sup>4</sup>School of Physical Education, Hunan Agricultural University, Changsha 410128, Hunan, China

<sup>5</sup>School of Physical Education, Jiujiang University, Jiujiang 332005, Jiangxi, China

Correspondence should be addressed to Chaobing Yan; 3150004@jju.edu.cn

Received 23 October 2020; Revised 5 December 2020; Accepted 5 January 2021; Published 3 February 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Xinwen Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The progress of social economy has created a better environment for the healthy development of young people, but the heavy schoolwork and life pressure have caused many students to ignore the scientific management of physical health. At this stage, people need a scientific physical health service system to help students understand their own health data, propose targeted exercise methods and health knowledge, and actively encourage and guide students to participate in physical exercise. The purpose of this article is to cultivate students' good self-exercise awareness and improve their physical fitness and health. To this end, this article has designed a smart health service system for young people. This article introduces the various service functions in the health management service system and explains in detail the entry, induction, and analysis of student physical health data in the system. The essence of the health intelligent service system is to provide students with targeted healthy exercise strategies through data analysis. This paper studies the health intervention plan of the health intelligent service system. From the experimental data, the improved particle swarm algorithm in this paper increases the effectiveness of the system in adolescent health data mining from 80.5% to 92.19%, which undoubtedly optimizes the system. It helps a lot.

## 1. Introduction

In recent years, the number of obese children in my country has increased, and the rate of adolescents' myopia has continued to increase. Physical fitness indicators such as vital capacity, speed, endurance, and explosive power have continued to decline. The physical health of adolescents has increasingly become a concern of the whole society. To ensure that young people have a healthy body, they must establish the concept of scientific fitness and have a certain degree of understanding of the basic physiology of the human body and exercise methods. Scientific fitness guidance can help young people improve the effects of exercise and effectively improve their physical fitness. In this era of the rapid development of computer technology, it is a very good choice to spread fitness information and fitness

guidance services through the Internet because the network can infinitely enlarge the capacity of stored information and provide more targeted fitness guidance for teenagers through big data intelligent calculation.

In foreign countries, many scholars have conducted research on the management system to improve the health of young people. Laurson cross-validated fitness standards related to Hungarian youth health. He collected data from 400 teenagers using treadmills to conduct aerobic capacity tests to determine peak oxygen consumption and bioelectrical impedance assessment data. Combined with the FitnessGram standard, young people are divided into healthy groups and groups in need of improvement, and logistic regression is used to estimate the probability of metabolic syndrome. Judging from the experimental data, he believes that 62.3% of the 400 research subjects are in a healthy state

of fitness. Of course, considering that there are only 400 people in the survey, this affects the persuasiveness of the data to a certain extent [1]. Welk hopes to establish an effective physical education program through the Internet to support and enhance the school's physical education program. He combined with FitnessGram, an American youth physical fitness assessment system, to provide unique insights on how to effectively provide large-scale school-based physical exercise. But on the whole, when he formulated the physical test data management, the design of the evaluation system was not perfect, and there is still a lot of room for improvement [2].

In recent years, the country has also begun to work on the organic integration of Internet technology and adolescent health management system. Wang believes that the overall health status of the youth in our country is on a downward trend at this stage, and a scientific health management system needs to be built for this. In his research, he elaborated on the information management plan of the leading thought of the system development and argued that testing is not the purpose, and collecting data should be to better judge the physical health of students and give further exercise suggestions. From the actual experimental results, the function of the system in terms of information feedback and supervision is not perfect [3].

The research of this article mainly starts with the following parts. First, this article introduces various technologies and methods in the health intelligent service system, including big data visualization technology, genetic algorithm and particle swarm algorithm in data mining algorithm, optimal sequence diagram method, and multiple methods of juvenile health literacy evaluation. Next, this article describes in detail the experimental development and database development of the health intelligent service system based on big data. Finally, this article conducts an in-depth analysis of the construction of the complex system of the adolescent physical health intelligent service system. The research focuses on the service functions in the health management service system, the entry, analysis, and calculation of health data, and the health intervention based on the health intelligent service system.

## 2. Technology and Methods in the Health Intelligent Service System

**2.1. Big Data Visualization.** The physical fitness test of young people is passed because these tests can reflect the physical health of the test subjects to a large extent. The big data visualization technology helps the test results to be displayed in a clearer way, allowing the young people to intuitively realize the health problems in their bodies and find the direction of fitness exercise [4]. The development of technologies such as big data storage, the Internet, and the Internet of Things has created better conditions for the processing of youth physical health data. Because physical health data has the characteristics of large data scale, multiple data types, low value of individual data, and fast data update, if you want to make an accurate comparison of the

physical health of adolescents, the accuracy and reliability of data processing are very high requirements [5].

**2.1.1. Distance-Based Visual Clustering Technology.** The visual layout technology proposed in this paper takes the parallel coordinate visual layout as the prototype, and the visual layout technology based on indicator health data enhances the visualization effect, making the visual analysis more clear and convenient for users [6]. The original parallel coordinate visualization layout does not enable users to quickly understand the distribution of data records as a whole. The distance-based visualization clustering method introduced in this section can make the data categories of different distribution characteristics in the data records clear [7]. Figure 1 is a schematic diagram of distance-based visual clustering.

In Figure 1,  $n$  represents the number of attribute axes, and  $d$  represents the distance between the two lines on the attribute axis, and they satisfy the following formula:

$$d_i = \|P_{ix} - P_{iy}\|. \quad (1)$$

Among them,  $p_{ix}, p_{iy}$  represents the coordinate position of the two polyline segments on the  $i$ -th attribute axis. Then the average distance  $\bar{d}$  formula and vertical coordinate calculation formula of two polyline segments satisfy the following:

$$\bar{d} = \frac{d_1 + d_2 + d_3 + \dots + d_n}{n}, \quad (2)$$

$$y'_p = F_a \cdot y_p + (1 - F_a) \cdot y_m.$$

In parallel coordinate visualization, each data record is connected to each other to form a polyline segment according to its position on each attribute. As the amount of data increases, the staggering phenomenon of parallel coordinate line segments becomes more and more obvious [8]. Set  $p_0, p_2$  to be the vertical coordinates of the two ends of the curve, and  $p_1$  to be the vertical coordinates of the quadratic interpolation point, then the intensity factor of the attraction within the class and the quadratic spline curve formula satisfy the following:

$$F_a = e^{-dh/\|y_m - y_p\|} \cdot \cos \alpha \cdot \Delta, \quad (3)$$

$$B(t) = (1 - t)^2 p_0 + 2t(1 - t)p_1 + t^2, \quad t \in [0, 1].$$

**2.1.2. Visual Aggregation Technology Based on Membership Degree.** The biggest advantage of scatter plots over other visualization layouts is that they can quickly get the distribution information of the data, such as the main distribution range of the data, the type of data, the characteristics of various data, and other information [9]. However, users can only get rough data distribution information through the points scattered in the graph, and the observation of the scattered point distribution in the three-dimensional scatter plot is not clear. The visual aggregation technology based on membership degree can better enable users to more quickly

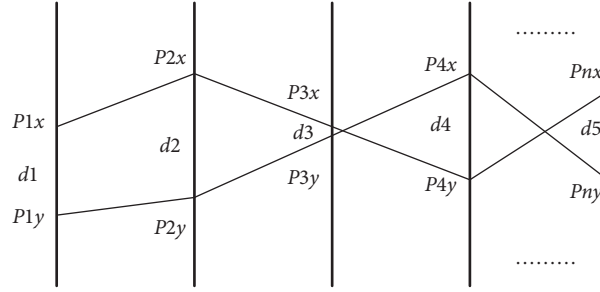


FIGURE 1: Schematic diagram of distance-based visual clustering.

obtain the distribution characteristics of the data records, thereby effectively enhancing the visualization effect of the scatter plot [10].

This method uses the objective function as the benchmark for clustering, and the objective function is as follows:

$$J = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \|x_j - c_i\|^2. \quad (4)$$

At the same time, there are constraints on the objective function:

$$\sum_{i=1}^c u_{ij} = 1, \quad j = 1, 2, 3, \dots, n, \quad (5)$$

where  $n$  is the number of scattered points,  $c$  is the number of classifications,  $u_{ij}^m$  is the membership degree of the  $j$ th scattered points relative to the  $i$ th category,  $m$  is the fuzzy weighting coefficient,  $c_i$  is the center coordinate of the  $i$ th category, and  $x_j$  is the  $j$ th scattered point coordinates. Then the distance between two points and  $c_i$  satisfy the following formula:

$$L = \|x_j - c_i\|, \quad (6)$$

$$c_i = \frac{\sum_{j=1}^n (x_j u_{ij}^m)}{\sum_{j=1}^n u_{ij}^m}.$$

It can be seen that in order to obtain  $c_i$ , the value of the membership matrix must be calculated first, then  $u_{ij}$  and the radius of the sphere satisfy the calculation formula:

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left( \|x_j - c_i\| / \|x_j - c_k\| \right)^{(2/(m-1))}}, \quad (7)$$

$$d_i = \frac{\sum_{k=1}^n \|x_{ik} - c_i\|}{n}.$$

**2.2. Data Mining Algorithm.** The main job of the health intelligent service system is to collect various physical information data of the youth group, judge the physical health of the students based on intelligent analysis, and give targeted exercise opinions [11]. However, the human body's health data is not only diverse in types but also very fast to update. In addition, there are still many data with low

reference value, which brings great adjustments to data analysis. Faced with this situation, it is necessary to find valuable information contained in a large amount of data through data mining technology to improve the efficiency and accuracy of data analysis [12]. Genetic algorithm (GA) and particle swarm optimization (PSO) are two very commonly used data mining algorithms.

**2.2.1. Genetic Algorithm.** The essence of a genetic algorithm is to simulate the genetic mechanism in nature and the theory of biological evolution. It has a very good performance when searching for the optimal solution randomly [13]. The genetic algorithm can be calculated as long as it has the information data of the objective function and will not be disturbed by spatial continuity. In addition, the genetic algorithm can also solve complex nonlinear problems and multidimensional space optimization problems, so it has a very wide range of applications in various fields [14, 15]. For a health management system based on big data, it is an excellent choice to analyze adolescent body data through genetic algorithms. Generally speaking, the genetic algorithm and the initial population can be formally described as follows:

$$\text{GA} = (P(0), N, I, S, g, p, f, t), \quad (8)$$

$$P(0) = (a_1(0), a_2(0), \dots, a_N(0)) \in I^N.$$

The basic operation steps of genetic algorithm are as follows: After clarifying the basic information of the objective function and variables of the problem, the variables can be coded, and the initial group  $\text{pop}_i(t)$  can be randomly generated according to the established coding method.

$$\text{pop}_i(t), \quad t = 1, 2, 3, \dots, N. \quad (9)$$

Individual fitness value is a concept used to evaluate and select individuals for various genetic operations. Generally speaking, fitness can reflect the excellence of each individual. The individual fitness value satisfies the following formula:

$$f_i = \text{fitness}(\text{pop}_i(t)). \quad (10)$$

When performing genetic algorithms, the value of individual fitness will directly affect operator selection in subsequent operations. Combining the value of the individual fitness function, it can be calculated that the selection probability of each individual satisfies the following formula:

$$P_i = \frac{f_i}{\sum_{i=1}^N f_i}, \quad i = 1, 2, 3, \dots, N. \quad (11)$$

Some excellent individuals are randomly selected from the original target group to complete genetic operations to form a new group of the next generation. The basic expression is as follows:

$$\text{newpop}(t+1) = \{\text{pop}_j(t), q = 1, 2, 3, \dots, N\}. \quad (12)$$

**2.2.2. Particle Swarm Algorithm.** Particle swarm algorithm is a kind of swarm intelligence algorithm. The inspiration of the swarm intelligence method comes from the research on the behavior of various biological groups in nature [16]. Compared with individual biological individuals, a large number of biological groups gathered together can more clearly reflect the behavioral characteristics of species. The steps of the particle swarm algorithm mainly include initializing particles, evaluating the fitness value of the particle by the fitness function, updating the particle swarm, checking whether it is terminated, and finally outputting the optimal solution.

Particle swarm optimization is also a population-based optimization algorithm like a genetic algorithm. It was originally used to solve continuous optimization problems and was specifically applied to function optimization and training neural network weights. The basic particle swarm algorithm satisfies the following formula:

$$\begin{aligned} v_i &= \omega \times v_i + c_1 \times \text{rand} \times (\text{pbest}_i - x_i) \\ &\quad + c_2 \times \text{rand} \times (\text{gbest}_i - x_i), \\ x_i &= x_i + v_i. \end{aligned} \quad (13)$$

Among them,  $\omega$  is the nonnegative inertia factor,  $c$  is the acceleration coefficient,  $x_i$  is the point in the D-dimensional space,  $v_i$  is the particle velocity, pbest is the individual extreme value, and gbest is the global extreme value. Since the parameter settings of the particle swarm algorithm have a great influence on the performance of the algorithm, it is mainly suitable for the optimization of continuous space functions. When applied to the optimization of health data in this paper, certain improvements and adjustments to the particle swarm algorithm are required [17].

**2.2.3. Particle Swarm Optimization.** At this stage, methods to improve the search accuracy of particle swarm optimization include dissipation theory, quantum update, chaotic update, etc. [18, 19]. Among them, quantum update and chaotic update have a good effect on the diversity of the population, and the dissipation theory can mainly improve the convergence speed and search ability of complex models, so these methods are beneficial to algorithm improvement. The following uses several standard mathematical model functions to test an improved real-number particle swarm algorithm and perform simple analysis of its performance.

$$f_1 = \sum_{i=1}^n \left( \sum_{j=1}^i x_j \right)^2. \quad (14)$$

Search space  $1 \leq x_j \leq 10, d = 20$ , global minimum  $f_1(x^*) = 0, x^* = 0, 0, \dots, 0$ .

$$f_2 = \sum_{i=1}^{n-1} \left( 100(x_{i+1} - x_i^2)^2 + (x_{i-1})^2 \right). \quad (15)$$

Search space  $1 \leq x_j \leq 10, d = 20$ , global minimum  $f_2(x^*) = 0, x^* = 1, 1, \dots, 1$ .

$$f_3 = \frac{1}{4000} \sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1. \quad (16)$$

Search space  $1 \leq x_j \leq 10, d = 20$ , global minimum  $f_3(x^*) = 0, x^* = 0, 0, \dots, 0$ .

$$f_4 = \sum_{i=1}^n \left( x_i^2 - 10 \cos(2\pi x_i) + 10 \right). \quad (17)$$

Search space  $1 \leq x_j \leq 10, d = 20$ , global minimum  $f_4(x^*) = 0, x^* = 0, 0, \dots, 0$ .

In order to improve the correlation between the test and the application effect of the physique test, each function model was randomly run 10 times, and the mean and variance of the running results were calculated. On the whole, quantum theory can effectively improve the search ability and search accuracy of the algorithm. For the effect of data mining in the health and intelligent service system, it is very necessary to improve the particle swarm algorithm [20].

**2.3. Sequence Diagram.** The principle of the optimal sequence diagram method is to compare and analyze the importance of each factor to the goal through the means of matrix illustration, and the purpose is to provide a basis for managers to make decisions [21, 22]. The optimal sequence diagram method is based on the number  $n$  of factors to be compared, and numbering the  $n$  factors. Let  $n$  be the number of comparison objects, and form a checkerboard pattern with  $n * n$  spaces for  $n$  comparison factors. When comparing factors in pairs, users need to compare the importance or pros and cons of the two and use numbers. To simply express the comparison result.

When judging the superiority or inferiority of each factor or the order of importance by using the superior sequence diagram method, it is necessary to check the validity of the results to ensure that the subjects are objective when comparing the various factors and will not appear because of the order in which the problems appear to be judgment deviation [23]. Assuming that the scores of each influencing factor are, respectively,  $A_1, A_2, \dots, A_i$ , then the weight of the  $i$ th influencing factor satisfies the following formula:

$$a_i = \frac{A_i}{\sum_{k=1}^{10} A_k}. \quad (18)$$

**2.4. Adolescent Health Literacy Evaluation Method.** The research of this paper is centered on the guidance service system of adolescent physical health based on big data. Since the construction of this system is based on the health needs of adolescents, it is necessary to comprehensively understand the health standards and scientific fitness needs of adolescents before proceeding with system development. To this end, this article has consulted a large amount of literature, combined with questionnaire surveys and interviews to collect basic personal information of some young people, their understanding of fitness knowledge, their acceptance of fitness guidance, and fitness needs. Since the satisfaction survey questionnaire used in this article is a qualitative factor questionnaire, it needs to pass the  $\alpha$  reliability coefficient method to accurately reflect the consistency between the scores in the scale [24]. The calculation formula meets the following:

$$\alpha = \frac{k}{k-1} \left( 1 - \frac{\sum s^2}{S_T^2} \right). \quad (19)$$

In addition, this article also uses SPSS statistical software and EXCEL to enter and analyze the surveyed data. Perform statistical processing on the data results to obtain the basis for quantitative analysis. At the same time, on the basis of the paper's interviews, questionnaire surveys and health management training camps, the collected data are analyzed and studied using logical thinking methods such as induction and deduction, analysis and synthesis, analogy, classification, and comparison [25].

### 3. Big Data-Based Health Intelligent Service System Experiment

**3.1. Experimental Background.** The successful development and promotion of adolescent health management system in some European and American countries can not only effectively improve the physical health of students but also have a very good effect in disease prevention. From a macroperspective, ensuring the physical and mental health of young people can also have a positive impact on promoting social harmony and economic stability. On the whole, this article can learn from the advantages of foreign health management systems and draw lessons from them to further promote the development of my country's physical health intelligent management system.

**3.2. Experimental System Design.** Adolescent fitness test and subsequent data processing and analysis are a series of tedious but meticulous work. The statistics of the data after the physical fitness test of all students, the results of data analysis, and the management of user data are not allowed to make mistakes. Therefore, this article needs to develop a system that can simplify the tedious work according to requirements. System administrators can use the system to conveniently manage information and data. Users can query their own information and physical test results through the system and obtain corresponding fitness program

recommendations. Figure 2 is a diagram of the functional modules of the juvenile physical health intelligent service system.

According to the system module functions in Figure 2, this article uses a lot of technologies in the development of the juvenile physical health intelligent service system, including Win8, Tomcat9, MySQL, Redis, and MAVEN; development tools include Eclipse Oxygen.Ia.Release, Navicat for Mysql and visual operation tools.

This article divides the use of the system into four parts: data administrators, general administrators, teachers, leaders, and ordinary students. After logging into the system, people with different identities have different permissions and available functions in the system. The module authority of the data administrator is to enter the data of the student's physical test into the system database and to manage and maintain the data on a regular basis. The general administrator can comprehensively manage the system. In addition to adding, deleting, and modifying data, it can also manage all user accounts, such as adding user IDs and modifying user passwords. The module permissions of teachers and leaders can query the physical test scores and various health scores of all students in their school as needed. Students can log in to the system with their own ID and password and inquire about their physical test scores and various health scores.

**3.3. Experimental Database Design.** The database is composed of many data tables. In the process of completing the database, it is necessary to write code to add configuration files required for reverse engineering of each data table. Among the more important data table diagrams in the database table include workbook, student mark, student info, body mass index, and record file info.

This study collects and analyzes the physical data of adolescents in various ways, and finally obtains the complete data requirements of the system. Based on a series of surveys, collections, and analyses, this article confirms the various information data required by the system, including student information data sheet, a list of student physical test scores, physical test index weighting tables for different grades, single index score sheets, bonus index score sheet, and application form for exemption from execution.

### 4. Construction of a Complex System Based on Big Data for the Intelligent Service System of Youth Physical Health

#### 4.1. System Analysis Module of Health Intelligent Service System

**4.1.1. Current Status of Adolescents' Health and the Construction of Management Service System.** When analyzing the status quo of adolescents' physical health, it can be considered from many aspects such as body shape, physical function, and physical fitness. Based on the survey data in recent years, the overall physical health of adolescents in my country has shown a downward trend. Figure 3 shows the



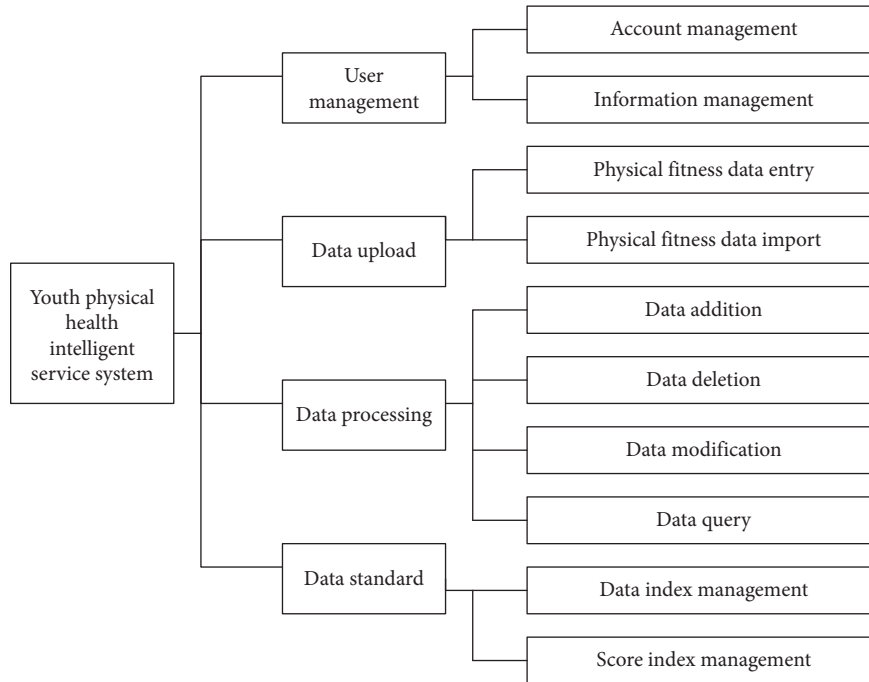


FIGURE 2: Functional module diagram of adolescent physical health intelligent service system.

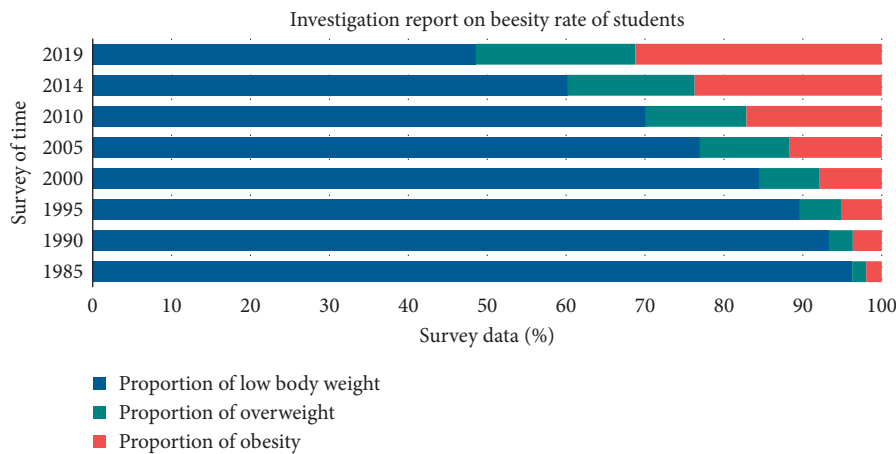


FIGURE 3: Statistics on the change in adolescent obesity rate from 1985 to 2019.

statistics on the changes in student obesity rates during the eight national student physical health surveys from 1985 to 2019.

It can be seen from Figure 3 that the proportion of malnutrition and low weight among Chinese adolescents has continued to decrease over the past 30 years, which corresponds to the continuous increase in the number of overweight and obese students. In the past ten years, the obesity rate of Chinese adolescents has nearly doubled. Even taking into account the factors of average height growth, these data are still worthy of attention.

There is no doubt about the positive effect of physical exercise on the physical health of young people. Maintaining three to four times a week of exercise is very helpful for students to improve their physical fitness and promote their

physical health. Figure 4 is a survey of the number of weekly exercises for students in a college.

It can be seen from Figure 4 that only 17.63% of students do physical exercises 3-4 times a week, and most students only do simple physical exercises during physical education classes. Therefore, the proportion of students who exercise 1-2 times a week is as high as 48.5%. In addition, there are even close to 30% of students who said they basically do not participate in physical exercise.

As an important part of student health management, the physical education department is responsible for helping students prevent health risks. At the current stage, physical education in schools often ignores the pertinence and intensity requirements for the classroom. In this context, the top priority is to continue to innovate ways to promote the

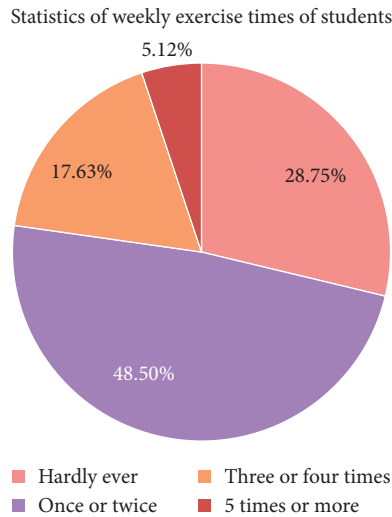


FIGURE 4: Investigation and statistics of students' weekly exercise times.

health of young people and to better provide systematic and scientific help and services for the health of young people.

At this stage, the school's physical examination simply records various indicators of the student's body and cannot effectively help students to correctly understand their own health status and arouses students' attention to strengthening their physical fitness. In the smart service system for youth physical health based on big data, based on the detection of student health information, the information data will be further summarized, analyzed, and evaluated, and targeted health guidance and intervention will be provided to help students improve their health. Comprehensive quality.

**4.1.2. Service Function in Health Management Service System.** In the development of a youth physical health intelligent service system, this article unites the school physical education department, physical test center, school hospital, mental health center, and food and nutrition science health resource sharing platform to unify these advantageous resource platforms to realize the sharing of data resources. By establishing individual health management files for young people, we can better manage and serve the health of students. Table 1 and Figure 5 show the percentage of data of the types of adolescent psychological counseling problems.

It can be seen from Figure 5 that the most common types of problems in adolescents' psychological distress are depression and anxiety, problems in personal growth (lack of self-confidence and trust, etc.), and stress in learning. It is worth noting that although learning pressure accounts for 68.1% of the troubled problems, only 23.7% of students are willing to go for psychological counseling.

To ensure the physical health of young people, it is very important to ensure that they have a healthy mind. While the pressure of social competition is increasing, the learning pressure that young people endure from an early age has also continued to increase. Excessive learning pressure is very detrimental to the physical and mental health of students.

Therefore, the health intelligence service system developed in this article cooperated with the Mental Health Center and added a psychological counseling section to the platform. In the online environment, young people can conduct psychological consultation anonymously and seek help from professional psychologists to reduce the probability of mental illness.

In the health intelligence service system, the school hospital undertakes various service responsibilities such as student body examination, physical fitness assessment, disease treatment, medical supervision, and health knowledge publicity. The prevention, diagnosis, and treatment of students' physical diseases in school hospitals can not only alleviate the pain caused by diseases to students but also reduce students' financial expenditure on medical treatment. School hospitals can help students strengthen their self-care awareness and improve their quality of life by carrying out regular health consultations, disease surveys, and promoting healthy living behaviors.

The Academy of Nutritional Sciences provides scientific diet guidance and scientific diet promotion services for students' health. In the platform, the system will provide corresponding scientific recipes based on student body data. For example, students who are underweight can add more protein and carbohydrates, and students who are overweight need to control their diet and reduce calorie intake. All in all, it is to help students develop healthy eating habits through a scientific diet.

#### 4.2. Data Analysis Module of Health Intelligent Service System

**4.2.1. Entry and Processing of Physical Health Data.** When entering adolescent physical health data in the health intelligence service system, the data administrator can choose the appropriate entry method according to the actual situation. The most traditional method is to collect paper reports of student test data, sort all the data, and enter them into the system database by class. In addition, there are now

TABLE 1: Proportion of types of adolescent psychological counseling problems.

|   | Depression and anxiety (%) | Suicide and other serious psychological problems (%) | Insomnia and eating disorders (%) | Personal growth problems (%) | Learning pressure (%) | Emotional problems (%) |
|---|----------------------------|--|-----------------------------------|------------------------------|-----------------------|------------------------|
| Proportion of troubled problems         | 76.2                       | 14.8   | 32.5                              | 72.6                         | 68.1                  | 35.2                   |
| Proportion of people willing to consult | 58.2                       | 51.6   | 40.3                              | 34.8                         | 23.7                  | 25.2                   |

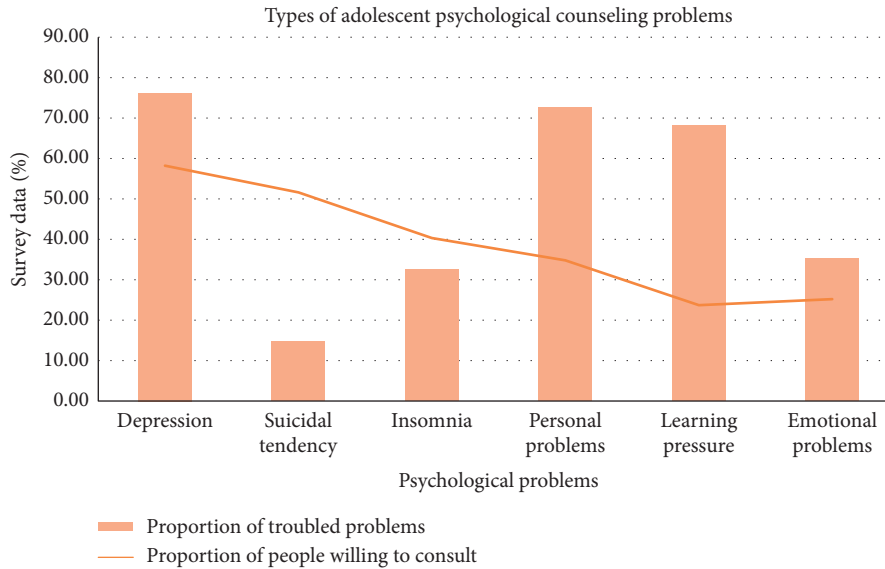


FIGURE 5: Types of adolescent psychological counseling problems.

many advanced testing instruments and systems that can automatically record data and generate excel forms. Administrators only need to add Excel forms into the health service system.

Using the health and wisdom service system, students can view various indicators in their physical fitness tests, and teachers can further understand the physical health of the student group by class. Table 2 is a single item score table for girls' body mass index (BMI). Figure 6 is a statistical chart of the body mass index of female students in a high school.

Compared with tables, statistical graphs can help teachers and leaders understand the overall health data of students more intuitively. It can be seen from Figure 6 that the overall body mass index of female students in the 21 classes of the high school from Grade One to Grade Three is in a normal state, but they are bound to be lighter and heavier students. Big data does not mean a general reference to data information. Choosing the correct data comparison indicators can help people further clarify the physical changes of students. Table 3 and Figure 7 show the body mass index of girls in the 7 classes in the second year of school. Figure 7 clearly shows the number of students who are lighter, heavier, and obese in each class. After the data is systematically summarized and sorted out, the effective calculation and analysis of the data health data can be ensured.

*4.2.2. Intelligent Calculation and Analysis of Physical Fitness Data.* In order to eliminate the interference items in a large amount of physical health data and correctly evaluate the health status of students, this paper uses intelligent data mining algorithms in the system. Figure 8 shows the change in accuracy of health data trait analysis before and after the improvement of the particle swarm algorithm.

It can be seen from Figure 8 that the improved particle swarm algorithm has increased its accuracy from 80.56% to 92.19% during the calculation, which is a good improvement. With the support of the improved particle swarm algorithm, the health intelligent service system based on big data can more accurately determine the health status of students from the physical fitness data of students and provide health intervention programs.

*4.3. Health Intervention Based on Health Intelligent Service System.* In the health intelligent service system, data collection, induction, and analysis are all for feature summary, and finally implement scientific health education intervention. In the smart health platform, people can not only understand the physical health of students through visual graphics but also harvest corresponding health education knowledge teaching and scientific exercise programs in the

TABLE 2: Female body mass index (BMI) single item score sheet.

| Grade       | Score | First grade | Second grade | Third grade | Senior one | Senior two | Senior three | University |
|-------------|-------|-------------|--------------|-------------|------------|------------|--------------|------------|
| Normal      | 100   | 14.8–21.7   | 15.3–22.1    | 16.0–22.6   | 16.5–22.7  | 16.9–23.1  | 17.1–23.3    | 17.2–23.8  |
| Underweight |       | ≤14.7       | ≤15.2        | ≤15.9       | ≤16.4      | ≤16.8      | ≤17.0        | ≤17.1      |
| High weight | 80    | 21.8–24.4   | 22.2–24.8    | 22.7–25.1   | 22.8–25.2  | 23.2–25.4  | 23.4–25.7    | 23.9–27.9  |
| Obesity     | 60    | ≥24.5       | ≥24.9        | ≥25.2       | ≥25.3      | ≥25.5      | ≥25.8        | ≥28.0      |

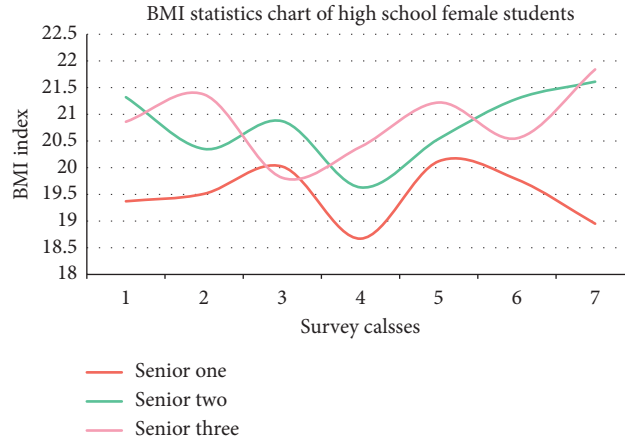


FIGURE 6: BMI statistics chart of high school female students.

TABLE 3: BMI situation of girls in 7 classes of high school.

|              | A  | B  | C  | D  | E  | F  | G  |
|--------------|----|----|----|----|----|----|----|
| Normal       | 31 | 16 | 27 | 21 | 26 | 19 | 22 |
| Under weight | 3  | 2  | 3  | 4  | 4  | 1  | 3  |
| High weight  | 5  | 2  | 2  | 3  | 6  | 5  | 4  |
| Obesity      | 2  | 1  | 3  | 2  | 4  | 3  | 3  |

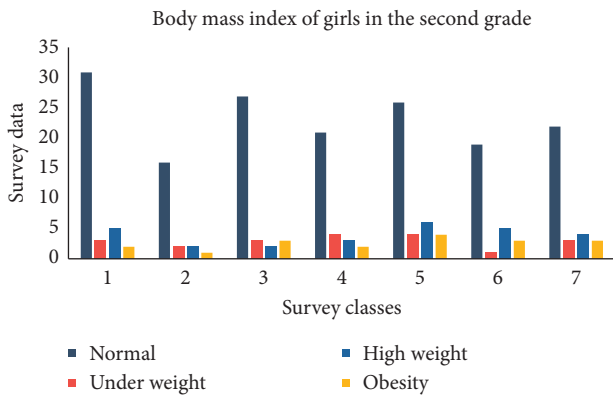


FIGURE 7: BMI situation of girls in 7 classes of high school.

health advice column. Table 4 shows the changes in various indicators after students receive scientific exercises.

It can be seen from Table 4 that the  $p$  values of all dimensions in the health survey summary table are less than 0.05, which indicates that the differences in the indicators of each dimension are significant after students listen to the health interventions given by the health smart service

system. Comparing the various dimensions of the health scale before and after the experiment, it is not difficult to see that the students' health in all aspects of the body has been significantly improved after scientific physical exercise. In the seven dimensions of the health survey summary table, the average value of each dimension has improved after the experiment than before the experiment. Especially in the dimensions of social function, emotional function, and physical pain. In terms of the overall health dimension, the average after the experiment also reached 86.31, which is a very high score, indicating that health management is of great help to the overall health of students. Figure 9 shows the changes in the average scores of female students after scientific exercise.

From Figure 9, after a period of exercise through the scientific intervention program recommended by the system, the students' performance in the physical test has been significantly improved. According to the comprehensive assessment standards, all students can exceed the passing line, and most students can get good grades and above. Figure 10 shows the survey data of students' satisfaction with the health intelligent service system.

It can be seen from Figure 10 that although traditional online fitness guidance also includes texts, pictures, videos, and other ways of guidance, it lacks understanding of students' physical conditions and cannot give targeted guidance. Therefore, students are satisfied with it. The degree is also relatively general. In contrast, the juvenile physical health intelligent service system built with big data has a significantly higher rate of praise among students.

Most youth associations lack a correct understanding of their changing physical conditions. At this time, scientific

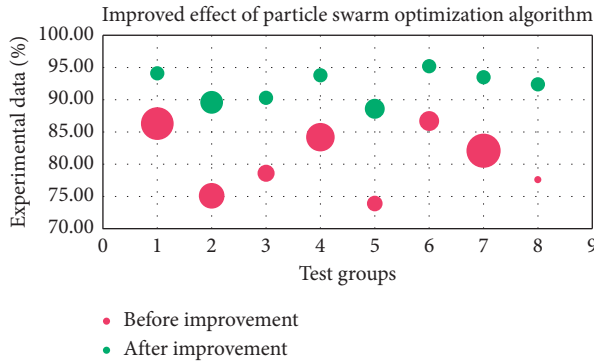


FIGURE 8: Optimization accuracy changes after particle swarm optimization improved.

TABLE 4: Statistical table of changes in various indicators after students' scientific exercise.

| Project                | Before exercise | After exercise | <i>t</i> | <i>p</i> |
|------------------------|-----------------|----------------|----------|----------|
| Physiological function | 68.22 ± 19.63   | 76.23 ± 19.98  | -15.63   | <0.001   |
| Body pain              | 72.53 ± 15.48   | 81.42 ± 13.51  | -17.88   | <0.001   |
| Social function        | 62.31 ± 25.88   | 81.84 ± 24.02  | -30.45   | <0.001   |
| Energy                 | 77.69 ± 14.21   | 84.21 ± 13.05  | -23.27   | <0.001   |
| Emotional function     | 62.76 ± 31.25   | 84.21 ± 24.83  | -21.20   | <0.001   |
| Mental health          | 83.21 ± 16.02   | 89.38 ± 11.37  | -15.32   | <0.001   |
| General health         | 74.23 ± 16.82   | 82.53 ± 16.01  | -33.21   | <0.001   |

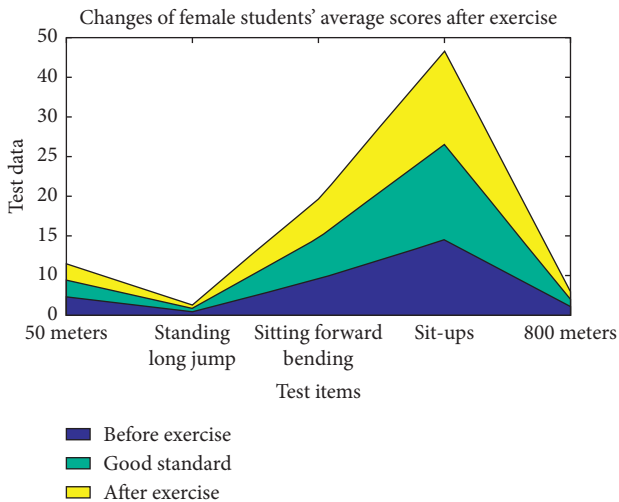


FIGURE 9: Change of students' average scores in physical examination.

physical health indicators are needed to provide students with reference to keep students more concerned about physical exercise. In the growth process of young people, both the body and the mind are in a state of constant maturity. The health service system based on big data is dedicated to accurately understanding the health status of

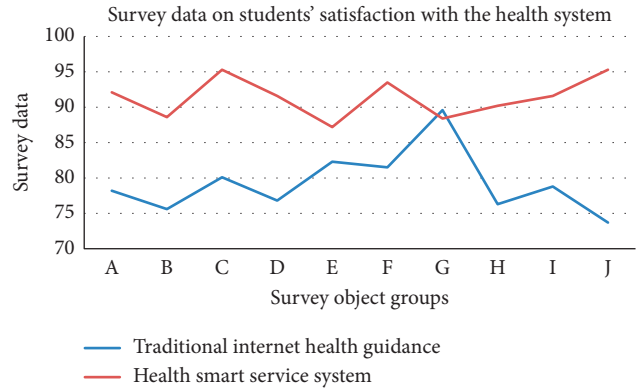


FIGURE 10: Survey data of students' satisfaction with the system.

each student and provides targeted health guidelines through intelligent analysis.

### 5. Conclusions

This article analyzes the system analysis module of the health intelligent service system. When analyzing the physical health of adolescents, it can be found that the overall physical health of adolescents in my country is on the decline. Therefore, helping students to form a correct understanding of physical exercise and choose appropriate exercise programs has become an important task of the physical health intelligent service system. When constructing the system, this article worked closely with the school's physical education department, physical test center, school hospital, mental health center, and the Academy of Food and Nutrition Sciences, and consolidated the data resources of each platform to establish the personal health of young people, managing files to provide scientific management and services for students' physical and mental health.

This article analyzes the data analysis module of the health intelligent service system. Based on the smart platform of big data, the most basic workflow is to provide powerful data support for people to make correct decisions through data collection, entry, induction, calculation, and analysis. In the growth stage, various physical data of young people are in the process of constant change. In order to accurately obtain the valuable part of a large amount of physical health data, this paper uses an improved particle swarm algorithm for data mining. From the experimental results, the improved algorithm increases the accuracy of data analysis from 80.5 to 92.19. Based on the data analysis results of the health intelligent service system, this article has conducted health education intervention and health exercise intervention for young people. After scientific exercise, the students' various physical indicators and physical test scores have been improved to a certain extent. On the whole, the physical intelligence service system can play a very good effect in promoting the physical and mental health of students.

Based on big data, this article has launched research on the construction of a youth physical health intelligent service

system. This article uses database analysis and processing technology to collect a large number of young people's physical health data, combined with intelligent algorithms to comprehensively analyze the data and provide scientific guidance and suggestions for students to perform physical exercises. Judging from the survey results, the system has achieved high levels of satisfaction among students, but on the whole, the platform still has many shortcomings. In future research, from the perspective of big data, we can realize the comprehensive management of youth physical health data, optimize the operation process of the system during work, and further improve the effect of data mining algorithms.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the Teaching and Research Project of University of Electronic Science and Technology of China, the Fundamental Research Funds for the Central Universities of China: Physical Education and Health Promotion (No. ZYGX2018007FRJH), and Building a Personalized and Customized System for Campus Physical Health and Creating a New Model of Physical Education (No. 2015XJYYB080).

## References

- [1] K. R. Laurson, P. F. Saint-Maurice, I. Karsai, and T. Csányi, "Cross-validation of FITNESSGRAM health-related fitness standards in Hungarian youth," *Research Quarterly for Exercise and Sport*, vol. 86, no. 1, pp. S13–S20, 2015.
- [2] G. J. Welk, Y. Bai, P. F. Saint-Maurice, K. Allums-Featherston, and N. Candelaria, "Design and evaluation of the NFL PLAY 60 FITNESSGRAM partnership project," *Research Quarterly for Exercise and Sport*, vol. 87, no. 1, pp. 1–13, 2016.
- [3] X. Q. Wang and Y. Y. Xu, "The exploration of building college students physical health management system," *Contemporary Sports Science and Technology*, vol. 6, no. 14, pp. 59–61, 2016.
- [4] G. Romanillos, M. Zaltz Austwick, D. Ettema, and J. De Kruijf, "Big data and cycling," *Transport Reviews*, vol. 36, no. 1, pp. 114–133, 2016.
- [5] R. A. Salam, J. K. Das, Z. S. Lassi, and Z. A. Bhutta, "Adolescent health interventions: conclusions, evidence gaps, and research priorities," *Journal of Adolescent Health*, vol. 59, no. 4, pp. S88–S92, 2016.
- [6] H. Al-Samarraie, S. M. Sarsam, and I. N. Umar, "Visual perception of multi-column-layout text: insight from repeated and non-repeated reading[J]," *Behaviour & Information Technology*, vol. 36, no. 1–3, pp. 75–84, 2016.
- [7] T. Itoh, A. Kumar, K. Klein et al., "High-dimensional data visualization by interactive construction of low-dimensional parallel coordinate plots," *Journal of Visual Languages & Computing*, vol. 43, no. 1, pp. 1–13, 2016.
- [8] T. Opach and J. K. Rød, "Augmenting the usability of parallel coordinate plot: the polyline glyphs," *Information Visualization*, vol. 17, no. 2, pp. 108–127, 2018.
- [9] H. Kim, J. Choo, H. Park, and A. Endert, "InterAxis: steering scatterplot axes via observation-level interaction," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 131–140, 2016.
- [10] L. Shao, T. Schleicher, M. Behrisch, T. Schreck, I. Sipiran, and D. A. Keim, "Guiding the exploration of scatter plot data using motif-based interest measures," *Journal of Visual Languages & Computing*, vol. 36, no. 1, pp. 1–12, 2016.
- [11] G. Luo, "Toward a progress indicator for machine learning model building and data mining algorithm execution," *Acm Sigkdd Explorations Newsletter*, vol. 19, no. 2, pp. 13–24, 2017.
- [12] X. Fei, G. Tian, S. Lima et al., "Research on data mining algorithm based on neural network and particle swarm optimization," *Journal of Intelligent & Fuzzy Systems*, vol. 35, no. 3, pp. 2921–2926, 2018.
- [13] Y. Tsujimura, M. Gen, and E. Kubota, "Solving job-shop scheduling problem with fuzzy processing time using genetic algorithm," *Journal of Japan Society for Fuzzy Theory & Systems*, vol. 7, no. 5, pp. 1073–1083, 2017.
- [14] U. Aickelin and K. Dowsland, "Exploiting problem structure in a genetic algorithm approach to a nurse rostering problem," *Journal of Scheduling*, vol. 3, no. 3, pp. 139–153, 2015.
- [15] M. Qiu, Z. Ming, J. Li, K. Gai, and Z. Zong, "Phase-change memory optimization for green cloud with genetic algorithm," *IEEE Transactions on Computers*, vol. 64, no. 12, pp. 3528–3540, 2015.
- [16] J. Sun, X. Wu, V. Palade et al., "Random drift particle swarm optimization algorithm: convergence analysis and parameter selection," *Machine Learning*, vol. 101, no. 1–3, pp. 345–376, 2015.
- [17] Y. Zhang, S. Wang, and G. Ji, "A comprehensive survey on particle swarm optimization algorithm and its applications," *Mathematical Problems in Engineering*, vol. 2015, no. 19, pp. 1–38, 2015.
- [18] K. M. Lee, T. T. Wu, and R. W. P. King, "Theory of an insulated linear antenna in a dissipative medium," *Radio Ence*, vol. 12, no. 2, pp. 195–203, 2016.
- [19] Z. Jiangbo and C. Ge, "Convergence rate of the asymmetric defluent-weisbuch dynamics," *Journal of Systems Ence and Complexity*, vol. 28, no. 004, pp. 773–787, 2015.
- [20] A. A. A. Esmine, R. A. Coelho, and S. Matwin, "A review on particle swarm optimization algorithm and its variants to clustering high-dimensional data," *Artificial Intelligence Review*, vol. 44, no. 1, pp. 23–45, 2015.
- [21] W. Hou, C. Shan, Y. Yu, P. Hu, and H. Zhang, "Modular platform optimization in conceptual vehicle body design via modified graph-based decomposition algorithm and cost-based priority method," *Structural and Multidisciplinary Optimization*, vol. 55, no. 6, pp. 2087–2097, 2017.
- [22] R. Xu, L. Zhang, N. Ge et al., "Schedulability analysis of graph-based real-time task model with precedence constraints," *International Journal of Software Engineering and Knowledge Engineering*, vol. 28, no. 11–12, pp. 1575–1595, 2018.
- [23] U. Laube and M. E. Nebel, "Maximum likelihood analysis of the ford-fulkerson method on special graphs," *Algorithmica*, vol. 74, no. 4, pp. 1224–1266, 2016.
- [24] M. J. Warrens and J. Matthijs, "A comparison of reliability coefficients for psychometric tests that consist of two parts," *Advances in Data Analysis and Classification*, vol. 10, no. 1, pp. 71–84, 2016.

- [25] J. F. Salgado, "Estimating coefficients of equivalence and stability for job performance ratings: the importance of controlling for transient error on criterion measurement," *International Journal of Selection and Assessment*, vol. 23, no. 1, pp. 37–44, 2015.

## Research Article

# Flood Simulation Analysis of the Biliu River Basin Based on the MIKE Model

Qi Liu <sup>1,2</sup>, Dianwu Wang <sup>3</sup>, Yulong Zhang,<sup>1</sup> and Li Wang<sup>4</sup>

<sup>1</sup>Shenyang Agriculture University, College of the Land and the Environment, Shenyang, Liaoning 110866, China

<sup>2</sup>Liaoning Agricultural College, Yingkou, Liaoning 115009, China

<sup>3</sup>Liaoning Provincial Water Resources Department, Shenyang, Liaoning 110003, China

<sup>4</sup>Fushun Vocational Technology Institute, Department of Chemical Engineering, Fushun, Liaoning 113122, China

Correspondence should be addressed to Dianwu Wang; 20131104@stu.syau.edu.cn

Received 14 September 2020; Revised 15 October 2020; Accepted 21 January 2021; Published 31 January 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Qi Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The Biliu River is the largest river in Dalian. The occurrence of floods and droughts in this basin has extremely important impacts on local industry, agriculture, and urban development. For a long time, the annual distribution of precipitation in the Biliu River Basin is extremely uneven, the river runoff varies greatly from year to year and season to year, floods and droughts occur frequently, and serious soil erosion results in fragile ecological environment and severe shortage of water resources. In this paper, the spatial and temporal changes of rainfall and runoff in the Biliu River Basin are studied through the coupling of the MIKE 11 model and the MIKE SHE model. The hydrological changes in the Biliu River Basin are simulated. The coupled model is verified by monthly runoff data from 1996 to 2015, and the simulation values are found to be true. The values match well. Based on the cyclical pattern of precipitation and runoff in the Biliu River Basin, the rainfall and runoff data in the Biliu River Basin from 2016 to 2030 are derived. The MIKE SHE/MIKE 11 coupling model is used to predict the Biliu River from 2016 to 2030. The results show that flood disasters are expected to occur in August 2020, July 2025, and July 2030, which can provide a basis for hydrological management in the Biliu River Basin.

## 1. Introduction

The Biliu River is the largest river in the south of Liaoning Province. The basin runs through Yingkou and Dalian. The flood and drought conditions in the basin directly affect the industrial and agricultural development and economic growth of the people in the two places. The Cocoon Hydrological Station is located in Taudaoling Village, Guiyunhua Township, Zhuanghe City, Liaoning Province. It has a longitude of 122° 33'E and a latitude of 40° 02'N. It was established in 1958. It is the longest station in the river basin and has the longest statistical period. The comprehensive hydrological monitoring station with the longest and most complete hydrological and meteorological data is the only provincial hydrological monitoring point in the Biliu River Basin. The Biliu Hydrological Station is located in Shuangta Town, Pulandian City, Liaoning Province, with 122° 28' east

longitude and 39° 49' north latitude. The station was established in 1986. The total length of 42 km from Cocoa Field Station to Biliu Reservoir is a relatively concentrated area for catchment of water in the upper and upper reaches of the Biliu River Basin. The rainfall during the flood season at each station exceeds 70% of the annual rainfall [1, 2]. The hydrological conditions of Hehe Reservoir can well reflect the flood and waterlogging rules of the Biliu River, which is of great significance to the flood forecasting of the Biliu River Basin [3].

With the deepening of research on natural phenomena, accurate description and scientific disclosure of rainfall-runoff patterns in the natural world through mathematical methods can help people better understand and predict the water cycle in nature and rationally and scientifically allocate and use water resources. This has become a hotspot in hydrology research [4–6]. The distributed hydrological



model is to digitize rainfall, evaporation, and runoff data through elevation to show the hydrological change process of runoff and sink, which has higher requirements for the accuracy of the data [7]. In terms of professional software, the Danish Institute of Water Resources and Water Environment has been a world leader [8]. The company's software has been verified in actual engineering and has been widely recognized in water resources research. The functions of the software cover production, sink, and flow, including land, rivers, lakes, and oceans. The spatial scope ranges from one dimension to three dimensions. The research area ranges from water environment to hydrodynamics, and the research business ranges from water resource assessment and management to surface water and groundwater [9, 10]. In combination, one-dimensional river networks, two-dimensional estuaries, and three-dimensional deep seas are involved. The Danish Institute of Water Resources and Water Environment is a professional research department dedicated to water resources and the water environment [11]. Their technology has always been a leader in the world [12]. Hydrology in France, the United Kingdom, and Denmark has developed the first hydrological model for small watersheds, SHE [13]. In order to solve the calculation relationship between different hydrological processes and the values of different processes at the same time, the ASHE Association splits and combines the components that make up the SHE model and initially realizes the coupling [14]. The Danish Institute of Water Resources (DHI) uses the same code to port all the important coupled components. Relevant scholars improved the MIKE SHE model through the techniques of MODFLOW and OPENMI [15, 16]. MIKE 11 is a widely used one-dimensional river canal hydrodynamic model. Related scholars used the MIKE 11 modeling system to manage water resources in Strymonas River and Kerkini Lake and achieved good results [17, 18]. MIKE SHE is widely used in various runoff simulations. The researchers took the Saliberta Basin in the middle of the Mahanadi River Basin as the research object and focused on the application of the physical-based distributed modeling system MIKE SHE with good results [19]. Relevant scholars use the spatial distribution model MIKE SHE SWET to analyze the impact of afforestation and tree species on hydrology in different watersheds and use MIKE 11/MIKE SHE to study regional water resources management [20, 21].

This paper uses the hydrometeorological data from southern Liaoning from 1978 to 2016, based on an in-depth analysis of the occurrence of rainfall, evaporation, and runoff and its impact on floods and droughts. The dynamic process of runoff drainage is numerically simulated, the characteristics of floods and droughts in the Biliu River Basin are systematically studied, and related mechanisms are discussed. Local suggestions for controlling natural disasters such as floods and droughts and strengthening ecological construction are proposed.

The rest of this article is organized as follows. Section 2 analyzes the collection and processing of data. The model is established in Section 3, and the model is verified and predicted in Section 4. Finally, Section 5 summarizes the full text.

## 2. Data Collection and Processing

According to the actual conditions of the Biliu River Basin, the channel conditions in the simulated section, and the ultimate goal of the study, several modules such as river runoff (OC), slope flow (OL), and unsaturated flow (UZ) were selected in the MIKE 11 model.

For all hydrological processes and selected modules of the MIKE 11 model, we collate the collected data, summarize the model range, terrain, rainfall, vegetation, and other data, establish SHE parameters, and fully prepare for the modeling. The range of the Biliu River Basin is shown in Figure 1.

*2.1. Terrain.* The topographic data come from the topographic map with elevation data downloaded from Ovi Map, including the water system map of the simulated watershed. The elevation data come from the geographic and national conditions monitoring cloud platform; through the adjustment of the coordinate system and the rasterization of ArcGIS, the corresponding topographic water system map is generated.

*2.2. Land Cover/Use.* The land cover/use data is from the historical statistics of the Ministry of Agriculture. The land cover type, soil type, and vegetation classification are all extracted from the observation data with a resolution of 500 m. The data projection is converted and merged with the MODIS Reprojection Tool (MRT). Based on this, a commonly used Gauss Kruger format is formed, which can facilitate data synchronization processing.

There are multiple classification systems for land cover. This study uses a common vegetation classification scheme for agriculture. Most of the Biliu River Basin is a rural area, and the land use has not changed much between 1978 and 2016, and it is relatively stable. Therefore, the land use type in 1999 was used for treatment and it was divided into forest land (coniferous forest, broad-leaved forest, and shrubs) accounted for 45.41%, arable land (all agricultural land) accounted for 42.45%, residential and industrial land accounted for 7.07%, garden land accounted for 1.66%, unused land accounted for 0.02%, and water area accounted for 3.39%. The land-use types in the Biliu River Basin are shown in Table 1.

*2.3. Soil Types.* Based on the local soil data query, the soil types and their parameters in the Biliu River Basin are shown in Table 2. It can be found from Table 2 that the saturated hydraulic conductivity of brown earth soil is the highest; the saturated moisture content of meadow soil is the highest; the field water holding capacity of paddy soil is the largest; in addition, the withered moisture content of paddy soil is also the biggest.

*2.4. Plant Foliar Index.* The leaf surface index (LAI) of each coverage type in this research process can be determined based on actual measurement results, and the relevant results are shown in Table 3.

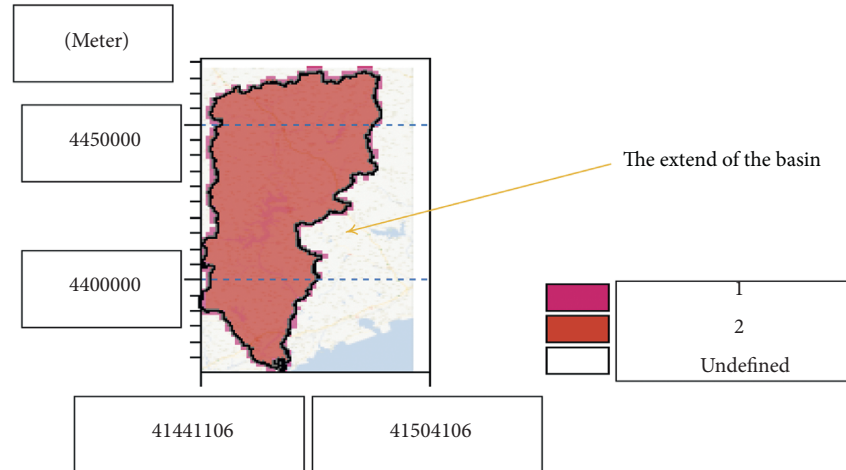


FIGURE 1: The extent of the Biliu River Basin.

TABLE 1: Land-use types in the Biliu River Basin.

| Soil use type             | Proportion (%) |
|---------------------------|----------------|
| Woodland                  | 45.41          |
| Arable land               | 42.45          |
| Garden                    | 1.66           |
| Waters                    | 3.39           |
| Urban and industrial land | 7.07           |
| Unused land               | 0.02           |
| Total                     | 100            |

2.5. *Others.* The four hydrological stations of the Biliu River are cocoon field, Xiaosongjiatun, Biliu Reservoir, and a tributary Yushi Reservoir. The annual average runoff is 612 million  $m^3$ .

Parameters related to rainfall, evaporation, and runoff were determined according to the statistics of the Hydrological Bureau. The meteorological data used in the calculation and analysis came from the Liaoning Meteorological Bureau. The location of the Biliu River Basin is shown in Figure 2.

There are obvious similarities between the average rainfall and annual rainfall distribution in the river basin during the flood season. The entire river basin gradually decreases from Dajiangtun in the north to Kuangdonggou in the southwest and Chengzitan in the south. All areas are gradually decreasing from the northeast to the southwest. The areas with increased rainfall are located in Dongsandaogou and Taipingzhuang in the northeast. The rainfall is greater than 560 mm, while the areas with lower rainfall are located in Bangladesh, in which the rainfall is less than 500 mm.

### 3. Model Establishment

3.1. *The Basis of the Model.* The runoff depth of each station does not change with time. The runoff depth of the Yushi Reservoir site has become smaller due to the commissioning of the Yushi Reservoir in 2003. At the same time, the Yushi Reservoir also affects the runoff depth of the mainstream cocoon field of the Biliu River and Xiaosongjiatun. Xiaosongjiatun Station is located in the lower

reaches of the Biliu River, and its runoff is greatly affected by the Biliu River Reservoir. In addition, the gradual decrease in rainfall over time is also the main reason for the depth of runoff. The blockage of river channels and the interception of vegetation will cause a decrease in runoff depth.

In the process of hydrological research, it is generally believed that a high-water year is a year when the flow is greater than the normal flow, while the opposite is true for a low-water year. The results of this paper show that the Biliu Reservoir has had more wet years than dry years in the past 40 years, while the results of other sites are opposite. The overall analysis shows that the ratio of the wet year to the dry year of the Biliu Reservoir is about 3 : 2, and the flood season is basically in line with the annual average, which shows that the regulation effect of the Biliu Reservoir is obvious.

The modeling process adopts the section from the mainstream cocoon field in the Biliu River Basin to the Biliu Reservoir. This section of the river has a large runoff and frequent floods. It plays a very important role in preventing floods in the Biliu River Basin. During the simulation, MIKE SHE was used to simulate the runoff, rainfall, and evapotranspiration fitting conditions at the cocoon field station, and MIKE 11 was used to simulate the water level change of the Biliu Reservoir.

Combined with the design data of the reservoir, the flood prevention alert water level is set. In this study, the Biliu Reservoir warning line 69.8 m designed by Liaoning Province Water Conservancy Survey and Design Institute in 1982 was used as an indicator of flood occurrence, and the maximum safe water level was 68.1 m. If the Biliu Reservoir water level reaches or exceeds the warning level, floods will occur and the reservoirs need to be flooded.

There are many factors that affect runoff changes, but rainfall is the most important factor. Analyzing the relationship between rainfall and runoff at various stations in the Biliu River Basin, it is found that the overall trend of annual runoff and rainfall is consistent. Secondly, the completion and commissioning of the Yushi Reservoir in 2003 greatly improved the hidden flood hazards and runoff conditions in

TABLE 2: Calculation of soil parameters.

| Numbering           | Withering water content (% vol) | Tianwen water holding capacity (% vol) | Saturated water content (% vol) | Saturated hydraulic conductivity (10 <sup>-6</sup> m/s) | Note                   |
|---------------------|---------------------------------|--|---------------------------------|---|------------------------|
| Brown soil          | 14.3                            | 27.1                                   | 43.8                            | 3.72  | Calibration parameters |
| Brown soil          | 14.8                            | 26.5                                   | 36.5                            | 4.13  |                        |
| Paddy soil          | 30.8                            | 43.1                                   | 42.3                            | 2.8   | Calibration parameters |
| Coastal saline soil | 14.8                            | 28.3                                   | 45.6                            | 3.68  |                        |
| Meadow soil         | 12.6                            | 28.5                                   | 46.1                            | 3.78  | Calibration parameters |

TABLE 3: Vegetation leaf area index in the Biliu River Basin.

| Types                           | January | February | March     | April   | May      | June     |
|---------------------------------|---------|----------|-----------|---------|----------|----------|
| Arable land                     | 0       | 0        | 0         | 0       | 1.5      | 2.0      |
| Woodland                        | 0       | 0        | 0         | 0       | 3        | 6        |
| Garden land                     | 0       | 0        | 0         | 0       | 0.25     | 0.50     |
| Waters city and industry unused | 0       | 0        | 0         | 0       | 0        | 0        |
|                                 | 0.5     | 0.5      | 0.5       | 0.5     | 0.5      | 0.5      |
| Types                           | July    | August   | September | October | November | December |
| Arable land                     | 3.5     | 4.6      | 4.5       | 3       | 0        | 0        |
| Woodland                        | 6       | 6        | 6         | 3       | 0        | 0        |
| Garden land                     | 2       | 2        | 2         | 2       | 0        | 0        |
| Waters city and industry unused | 0       | 0        | 0         | 0       | 0        | 0        |
|                                 | 0.5     | 0.5      | 0.5       | 0.5     | 0.5      | 0.5      |

the Biliu River Basin, and the design standards of the Biliu Reservoir have been continuously improved. Third, the improvement of the river course reduces the leakage loss of the river course, so that the impact of runoff is not obvious when the annual rainfall continues to decrease.

There is a total of 42 km from the cocoon field to the Biliu Reservoir. During the modeling process, sections were established at Cocoon Station, 40.5 km from Cocoon Station, 41.99 km from Cocoon Station, and 42 km from Cocoon Station to ease the water flow. A weir was established at 41.999 km from the cocoon field station to control the water level. When the water level exceeds 89.2 m, the weir will automatically prevent flooding until it reaches a safe water level of 68.1 m.

Runoff concentration is to synthesize the monthly runoff by the vector method, and the ratio of the resulting composite amount to the annual runoff can be used to describe the concentration level of the runoff during the year. The concentration period refers to the orientation of this composite vector, which describes the month corresponding to the center of gravity of the annual runoff concentration. In the process of mathematical analysis, it is generally described by the tangent angle of the ratio of the component sum.

For the convenience of research, the Biliu Reservoir was generalized into a part of the river during the research. The river below the reservoir was set at a constant water level of 30 m and the natural slope drop ratio of the river was 2‰.

In this study, the 5-year moving average method was used to select typical stations on the Biliu River, namely, the cocoon site, Biliu Reservoir, Xiaosongjiatun, and Yushi Reservoir, to analyze the change trend of its runoff. Then, the MK method was used to analyze the cocoon site and Biliu River. The MK method is used to test the sudden change of the rainfall trend. The MK method is a nonparametric test method. It does not need to analyze the distribution of the sample. The impact of abnormal values is very small. It does not need to be considered. It has high applicability to various types of variables. The accuracy of the result obtained is high, and the calculation is simple. The M-K method is widely used in hydrological research and meteorological research trend checking.

For the climate sequence  $x_i$ , at time  $i$ ,  $i = 1, 2, \dots, n-1$ , there are

$$m_j = \begin{cases} 1, & x_j > x_i, \\ 0, & \text{others,} \end{cases} \quad (1)$$

$$r_i = \sum_{j=i+1}^n m_j, \quad (2)$$

$$Z = \frac{4 \sum_{i=1}^{n-1} r_i}{n(n-1)} - 1. \quad (3)$$

We set the significance level to  $\alpha = 0.05$ , and then the criterion is

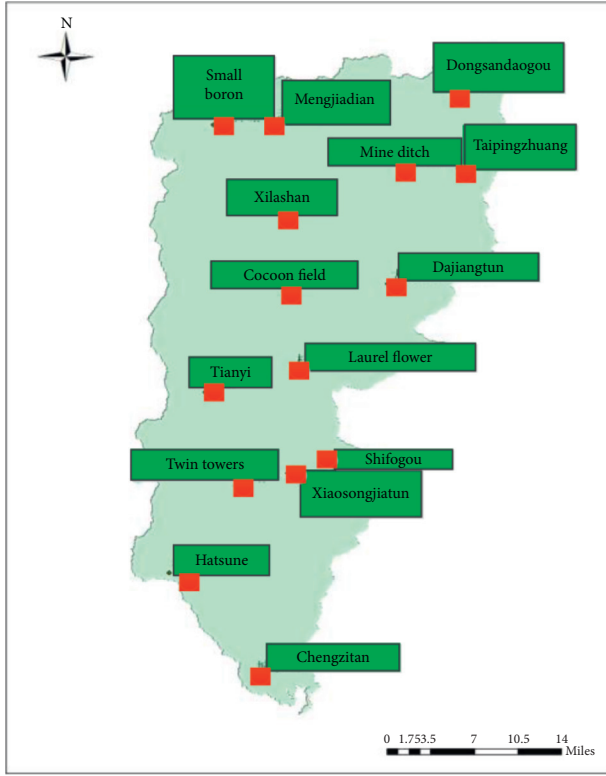


FIGURE 2: Survey station map of the Biliu River Basin.

$$Z_{\alpha} = 1.96 \times \left[ \frac{4n + 10}{9n(n-1)} \right]^{(1/2)}. \quad (4)$$

Suppose that the sequence is  $x_1, x_2, \dots, x_n$ ,  $S_k$  represents the cumulative number of the  $i$ -th sample  $x_i > x_j$  ( $1 \leq j \leq i$ ), and the defined statistics are

$$S_k = \sum_{i=1}^k r_i, r_j = \begin{cases} 1, & x_i > x_j, \\ 0, & x_i \leq x_j. \end{cases} \quad (5)$$

**3.2. Model Establishment.** The MIKE SHE model runoff simulation is performed by the Saint-Venant equations, where the momentum equation is simulated by the following equations:

$$\frac{\partial h}{\partial t} + \frac{\partial(uh)}{\partial x_i} + \frac{\partial(vh)}{\partial x_j} = q, \quad (6)$$

$$\frac{\partial(h)}{\partial x_i} = S_{oxi} - S_{fxi}, \quad (7)$$

$$\frac{\partial(h)}{\partial x_j} = S_{oxj} - S_{fxj}. \quad (8)$$

In the formula,  $h(x_i, y_j)$  represents the surface water depth (m),  $(x_i, y_j)$  represents the space coordinates,  $t$  is time (s),  $u$  and  $v$  represent the surface runoff velocity (m/s), and  $q$  represents the horizontal direction of the source-sink term

of inflow per unit area ( $\text{m}^3/\text{s m}^2$ ),  $S_{oi}$  and  $S_{oj}(x_i, y_j)$  represent the ground slope in the directions of  $x_i$  and  $y_j$ , and  $S_{fi}$  and  $S_{fj}(x_i, y_j)$  represent downhill in the directions of  $x_i$  and  $y_j$ .

The rainfall and runoff of the Biliu River Basin were simulated, and MIKE 11 and MIKE SHE were coupled for flood forecast research.

**3.2.1. MIKE 11 Model.** We establish a river network file based on the DEM data in the Biliu River Basin as required. The cross-section data are converted into a .shp file and entered into the MIKE 11 model to adjust the hydrodynamic parameter data.

The cross-section data simulated by the MIKE 11 model are derived from the generalization of the reservoir channel design and actual topographical data. The amount of water comes from the runoff generated by rainfall at the confluence of the cocoon field station. The cocoon field station is located 59 km from the river source; the lower boundary is 101 km from the river source. The river basin in this section is located in the middle of the Biliu River Basin. The concentrations of river runoff and large drop are of great significance for studying the drought and flood laws and forecasting in the Biliu River Basin. The height difference between the upper and lower boundaries will be calculated from the average ratio of the river channel and the actual relative height difference. Hydrodynamic parameters are set by default.

**3.2.2. MIKE SHE Model.** In the model calculation area, the mainstream of the Biliu River Basin was selected. When the grid was set, the grid size was determined to be 1 km. The target area was divided to determine  $50 \times 100$  units. In the course of this research, for the convenience of preprocessing and supplementation, all of them were uniformly transformed into the Xi'an 80 coordinate system. The rainfall data can be processed and determined based on the measured values of various stations in this area. The obtained results are processed appropriately to form the corresponding time series file (\* .dfs0); the evaporation data is determined according to local meteorological statistics. The Penman formula is calculated on the obtained raw data, and the resulting ET0 value is used to generate a \* .dfs0 file. In the model, the rainfall data and evaporation data are set separately by Tyson polygons. Thiessen polygon is a set of continuous polygons composed of vertical bisectors connecting two adjacent points. The distance from any point in a Thiessen polygon to the control points constituting the polygon is less than the distance to the control points of other polygons.

**3.2.3. Coupling of MIKE SHE/MIKE 11.** The water movement module of MIKE SHE covers the main processes of the hydrological cycle. It can study the hydrological cycle processes at different scales. It describes the hydrological cycle process of the runoff through evapotranspiration, slope flow, rivers and lakes, unsaturated flow, and saturated flow: evapotranspiration (ET) related factors mainly include

interception, soil, and water surface evaporation; surface runoff is described according to slope flow (OL), etc.; rivers and lakes (OC) are obtained by MIKE simulation analysis; and the correlation of reservoirs and dams is also calculated. However, the simulation of the hydrodynamic part of the MIKE SHE channel is insufficient and needs to be coupled with MIKE 11. Building a comprehensive MIKE SHE/MIKE 11 model includes the following basic steps.

First, we set the corresponding MIKE 11 hydrodynamic independent model and then test and analyze, under the possible conditions through the specified flow-related parameters to make a preliminary determination. Then, we establish the NIKE SHE model, the modules of which mainly include surface flow and saturated band, unsaturated band, and so on. Hydrodynamic simulation (1) and hydrodynamic simulation (2) are shown in Figures 3 and 4, respectively.

The branch river is defined as the connection between the two models, and the two are coupled. The lake submodule OC is used to import the simulation files related to the previous model. In this simulation system, the two models are simultaneously simulated and analyzed, and the data of the two is processed by the switch based on the shared storage space. MIKE 11 calculated the corresponding runoff water level, transferred the results to the latter model, and then compared the analysis results with the surface terrain information stored in the model to determine the remaining water volume. It can be inferred that the influential factors of water exchange between the two are mainly the corresponding evaporation and infiltration of surface water and the corresponding river exchange. Finally, the calculated water flow is exchanged with the first model through the sink source term of the Saint-Venant equation.

During the MIKE SHE model input process, the corresponding input data content is determined based on the corresponding modules and problem-solving requirements of the simulated hydrological process. The basic parameters include the model range, terrain, and rainfall. These data are indispensable in various MIKE SHE models. Based on the research area, a grid size of 1 km was set during the research in this paper. The target area was partitioned to determine  $50 \times 100$  cells. In the course of this research, for the convenience of preprocessing and supplementation, all data were converted and uniformly converted to the Xi'an 80 coordinate system.

We use MIKE 11 to simulate the river from the cocoon field in the Biliu River Basin to the Biliu Reservoir. According to the needs, the DEM data in the Biliu River Basin is used to create a river network file, and the cross-section data are converted into a .shp file to enter the MIKE 11 model to adjust the hydrodynamic parameter data.

The cross-section data simulated by the MIKE 11 model are derived from the generalization of the reservoir

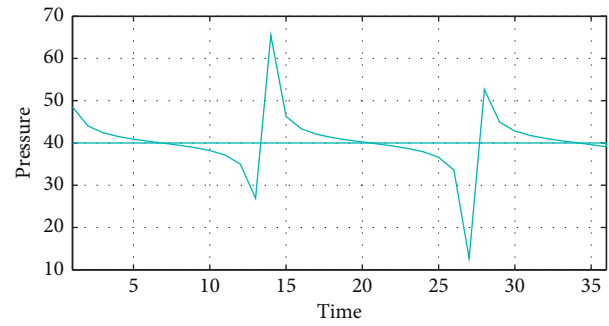


FIGURE 3: Hydrodynamic simulation (1).

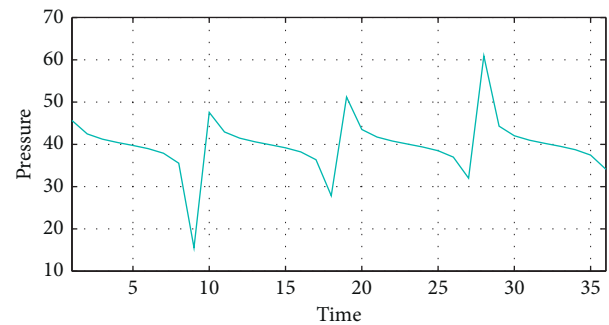


FIGURE 4: Hydrodynamic simulation (2).

channel design and actual topographical data. The amount of water comes from the runoff generated by rainfall at the confluence of the cocoon field station. The cocoon field station is located 59 km from the river source; the lower boundary is located 101 km from the river source; the river basin in this section is located in the middle of the Biliu River Basin. The concentrations of river runoff and large drop are of great significance for studying the drought and flood laws and forecasting in the Biliu River Basin. The height difference between the upper and lower boundaries will be calculated from the average ratio of the river channel and the actual relative height difference. Hydrodynamic parameters are set by default. The coupling between MIKE SHE and MIKE 11 is shown in Figure 5.

#### 4. Model Verification and Prediction

In order to verify the model effect, the adjustment of MIKE SHE model parameters should be as small as possible. In the MIKE SHE/MIKE 11 coupling model, the setting of parameters affects both the MIKE SHE and MIKE 11 models, and they are uniform in the river network. The correlation level between the measured value and the simulated value is

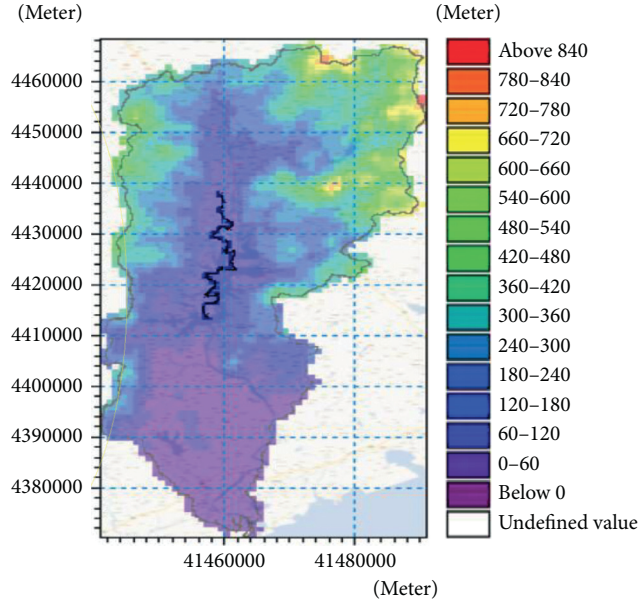


FIGURE 5: Coupled river pattern of MIKE 11/MIKE SHE.

described by the correlation coefficient, and the fitting degree of the two is evaluated by the Nash coefficient (Nse).

$$R = \frac{\sum_t (\text{Calc}_{i,t} - \overline{\text{Calc}_i}) \cdot (\text{Obs}_{i,t} - \overline{\text{Obs}_i})}{\sqrt{\sum_t (\text{Calc}_{i,t} - \overline{\text{Calc}_i})^2 \cdot \sum_t (\text{Obs}_{i,t} - \overline{\text{Obs}_i})^2}} \quad (9)$$

$$\text{Nse} = 1 - \frac{\sum_t (\text{Obs}_{i,t} - \text{Calc}_{i,t})^2}{\sum_t (\text{Obs}_{i,t} - \overline{\text{Obs}_i})^2} \quad (10)$$

Among them,  $\text{Obs}_i$ ,  $\text{Calc}_i$ ,  $\overline{\text{Obs}_i}$ , and  $\overline{\text{Calc}_i}$  are the measured and calculated values and their average values at  $i$ , respectively. The value range of Nse is negative infinity to 1. The larger the parameter is, the better the simulation effect can be judged to be, and the credibility requirements are met. If Nse is close to 0, it can reflect that the obtained result is close to the observation mean; that is, credibility is in the acceptance interval. If it is small, the model is untrustworthy. Relevant experience shows that if Nse is greater than 0.75, the simulation effect can be judged to be good; within the range of 0.36 to 0.75, it can be considered basically satisfactory; if it is less than 0.36, the requirements are not satisfied. Generally, when  $R > 0.8$ , it is considered that the level of linear correlation is high.

**4.1. Calibration of Parameters.** The MIKE SHE/MIKE 11 coupling model uses the data of the cocoon field from 1986 to 1995 for the parameter calibration. The calibration results are shown in Table 4. The simulation results of the runoff process before and after calibration are shown in Figures 6 and 7.

The correlation coefficient before calibration is  $R = 0.751294$ , and the Nash index is  $\text{Nse} = 0.531545$ ; the correlation coefficient after calibration is  $R = 0.94583$ , and the Nash index is  $\text{Nse} = 0.71379$ . The simulated and observed values achieve a good fitting effect.

**4.2. Model Verification.** The MIKE SHE/MIKE 11 coupling model was verified using two consecutive 10-year rainfall and runoff data from the cocoon field from 1996 to 2005 and 2006 to 2015. The results are shown in Figures 8 and 9, respectively. Time series verification results are from 1996 to 2005, correlation coefficient is  $R = 0.98818$ , and Nash index is  $\text{Nse} = 0.918427$ ; time series verification results are from 2006 to 2015, correlation coefficient is  $R = 0.964802$ , and Nash index is  $\text{Nse} = 0.895926$ . It shows that the simulated values and observed values of the two periods are in good agreement whether from the peak flow or the observation time.

Figures 10 and 11 show the changes in the water level at the Biliu Reservoir Station for two consecutive years (1996–2005 and 2006–2015). It can be found that, in the 20 consecutive years from 1996 to 2015, the water level of the Biliu Reservoir exceeded the warning level by 69.8 m: July 1995, July 1996, August 1997, August 1998, and August 2001, July 2005, August 2006, August 2010, August 2011, July 2012, and July 2013, these time nodes coincide with the historical changes in the water level of the Biliu Reservoir, indicating that the water level simulation results of the river reservoir can well reflect the real water level change of the reservoir. At this time, the reservoir needs to be discharged to prevent floods.

**4.3. Application of the Model and Flood Disaster Prediction.** The MIKE SHE/MIKE 11 coupling model was calibrated and verified by the rainfall, evaporation, and runoff data from the cocoon field from 1985 to 2015. It is found that the verification model is in good agreement with the real effect, and the water level changes of the Biliu Reservoir obtained from the simulation are basically consistent with the historical data of the Biliu Reservoir. This shows that the application of this model to predict flood disasters in the Biliu River Basin has a good effect.

TABLE 4: Parameter rate determination result.

| Parameter name                          | Calibration results   | Parameter unit  |
|---|-----------------------|-----------------|
| Evapotranspiration (ET)                 | 1.665                 | mm/d            |
| Slope roughness ( $M$ )                 | 5.0                   | $m^{1/3}s^{-1}$ |
| Initial water depth                     | -0.5                  | m               |
| Channel roughness (RM)                  | 5.0                   | $m^{1/3}s^{-1}$ |
| <i>Saturated hydraulic conductivity</i> |                       |                 |
| Brown soils                             | $4.13 \times 10^{-6}$ | $ms^{-1}$       |
| Brown earth                             | $3.72 \times 10^{-6}$ | $ms^{-1}$       |
| Coastal solonchaks                      | $3.68 \times 10^{-6}$ | $ms^{-1}$       |
| Meadow soils                            | $4.68 \times 10^{-6}$ | $ms^{-1}$       |
| Paddy soils                             | $2.80 \times 10^{-6}$ | $ms^{-1}$       |

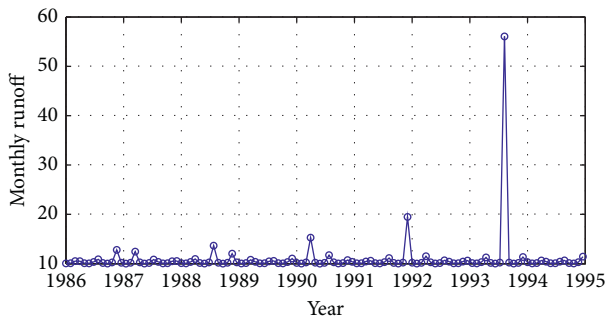


FIGURE 6: Simulation results of monthly runoff in Jianchang field (1986~1995).

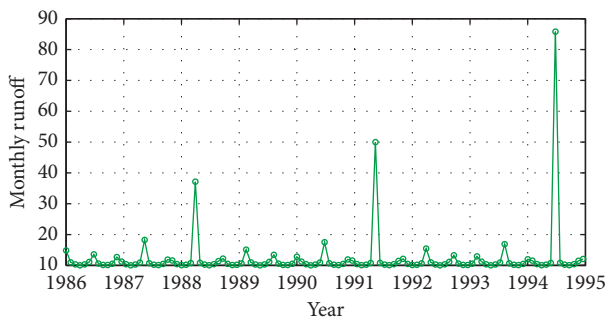


FIGURE 7: Simulated results of monthly runoff in the cocoon field after calibration (1986~1995).

Based on the cyclical pattern of precipitation and runoff in the Biliu River Basin from time to time, the precipitation and runoff data of the Biliu River Basin from 2016 to 2030 are derived. The MIKE SHE/MIKE 11 coupling model was used to predict flood disasters in the Biliu River Basin from 2016 to 2030.

Figure 12 shows the results of the monthly runoff simulation through the cocoon field. It can be seen that the correlation coefficient of the correlation coefficient between the derived value and the simulated value is  $R = 0.965588$ , the Nash index is  $Nse = 0.920947$ , and the fitting effect is very good. Figure 13 shows the water level change of the Biliu Reservoir. It can be seen that flood disasters are expected to occur in August 2020, July 2025, and July 2030.

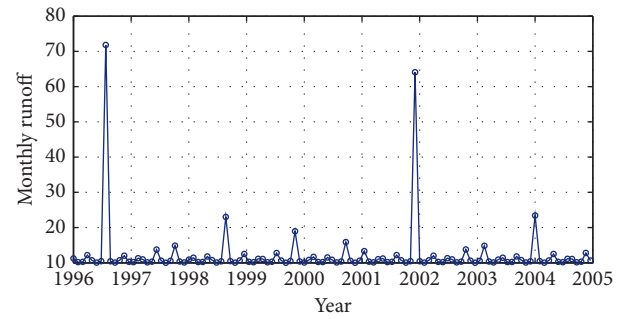


FIGURE 8: Simulation results of monthly runoff in Jianchang field (1996~2005).

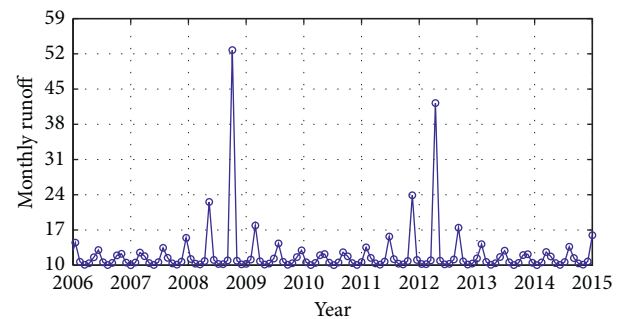


FIGURE 9: Simulation results of monthly runoff in Jianchang field (2006~2015).

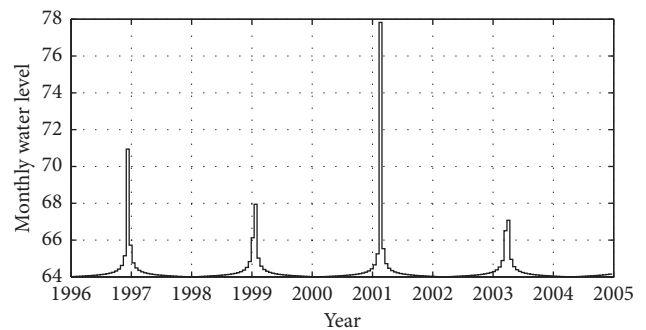


FIGURE 10: Simulation results of monthly water level in the Biliu River Reservoir (1996~2005).

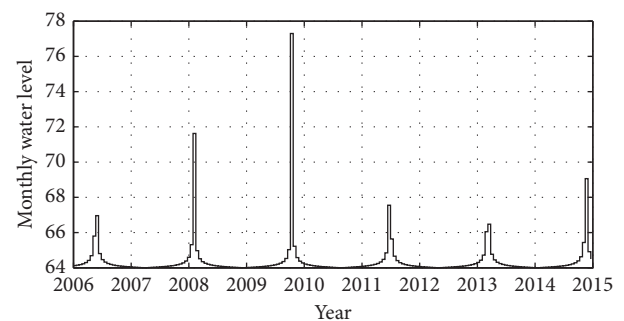


FIGURE 11: Simulation results of monthly water level in the Biliu River Reservoir (2006~2015).

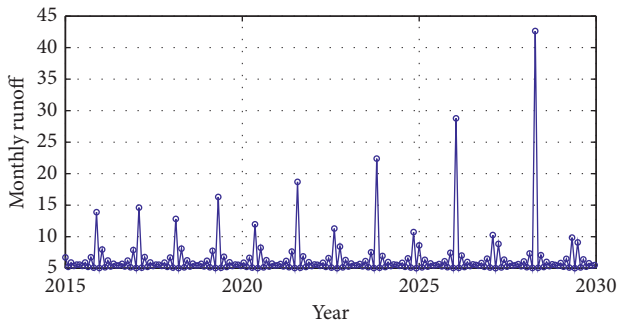


FIGURE 12: Simulation results of monthly runoff in Jianchang field (2005~2015).

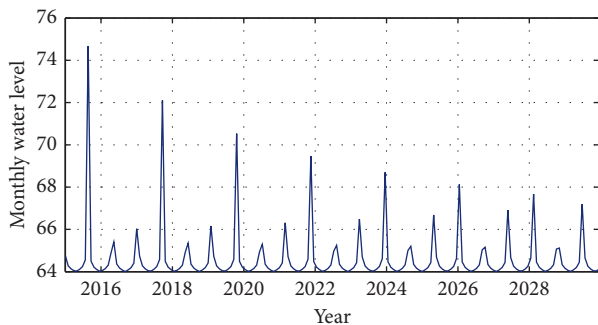


FIGURE 13: Simulation results of monthly water level in the Biliu Reservoir (2005~2015).

## 5. Conclusion

For the section of the Biliu River Basin from the cocoon field to the Biliu Reservoir, this paper uses the MIKE 11 model to create a one-dimensional hydrodynamic model. Through various terrain indicators, the MIKE SHE model is established. The coupling of MIKE 11 model and MIKE SHE model is used to study the temporal and spatial changes of rainfall and runoff in this basin and to simulate the hydrological changes in Biliu River Basin. Based on the monthly runoff data of Biliu River Basin from 1986 to 1995, the parameters were calibrated, and the model was given a reasonable parameter scheme. The model was verified through monthly runoff data from 1996 to 2015, and the simulation values were found to be in good agreement with the real values, which ensured the applicability of the model in the study area. Based on the cyclical pattern of precipitation and runoff in the Biliu River Basin from time to time, the precipitation and runoff data of the Biliu River Basin from 2016 to 2030 are derived. The MIKE SHE/MIKE 11 coupling model was used to predict flood disasters in the Biliu River Basin from 2016 to 2030. The results show that flood disasters are expected in August 2020, July 2025, and July 2030.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] T. A. Nigussie and A. Altunkaynak, "Modeling the effect of urbanization on flood risk in Ayamama Watershed, Istanbul, Turkey, using the MIKE 21 FM model," *Natural Hazards*, vol. 99, no. 2, pp. 1031–1047, 2019.
- [2] M. Kumar, "Sensitivity analysis of Manning's roughness coefficient in MIKE 11 model for Mahanadi delta region," *Annals of Plant and Soil Research*, vol. 21, no. 1, pp. 62–66, 2019.
- [3] H. X. Do, F. Zhao, S. Westra et al., "Historical and future changes in global flood magnitude—evidence from a model-observation investigation," *Hydrology and Earth System Sciences*, vol. 24, no. 3, pp. 1543–1564, 2020.
- [4] W. Li, K. Lin, T. Zhao et al., "Risk assessment and sensitivity analysis of flash floods in ungauged basins using coupled hydrologic and hydrodynamic models," *Journal of Hydrology*, vol. 572, pp. 108–120, 2019.
- [5] S. Shrestha, S. Cherng, A. N. Hill et al., "Impact and effectiveness of state-level tuberculosis interventions in California, Florida, New York, and Texas: a model-based analysis," *American Journal of Epidemiology*, vol. 188, no. 9, pp. 1733–1741, 2019.
- [6] W. Zhang, J. Li, Y. Chen et al., "A surrogate-based optimization design and uncertainty analysis for urban flood mitigation," *Water Resources Management*, vol. 33, no. 12, pp. 4201–4214, 2019.
- [7] E. Yildirim and I. Demir, "An integrated web framework for HAZUS-MH flood loss estimation analysis," *Natural Hazards*, vol. 99, no. 1, pp. 275–286, 2019.
- [8] J. Pinos, L. Timbe, and E. Timbe, "Evaluation of 1D hydraulic models for the simulation of mountain fluvial floods: a case study of the Santa Bárbara River in Ecuador," *Water Practice and Technology*, vol. 14, no. 2, pp. 341–354, 2019.
- [9] R. Wang, Y. Yuan, H. Yen et al., "A review of pesticide fate and transport simulation at watershed level using SWAT: current status and research concerns," *Science of The Total Environment*, vol. 669, pp. 512–526, 2019.
- [10] F. Ehmele and M. Kunz, "Flood-related extreme precipitation in southwestern Germany: development of a two-dimensional stochastic precipitation model," *Hydrology and Earth System Sciences*, vol. 23, no. 2, pp. 1083–1102, 2019.
- [11] M. L. Follum, R. Vera, A. A. Tavakoly et al., "Improved accuracy and efficiency of flood inundation mapping of low-, medium-, and high-flow events using the AutoRoute model," *Natural Hazards and Earth System Sciences*, vol. 20, no. 2, pp. 625–641, 2020.
- [12] J. L. Gutenson, A. A. Tavakoly, M. D. Wahl et al., "Comparison of generalized non-data-driven lake and reservoir routing models for global-scale hydrologic forecasting of reservoir outflow at diurnal time steps," *Hydrology and Earth System Sciences*, vol. 24, no. 5, pp. 2711–2729, 2020.
- [13] V. Ganti, M. P. Lamb, and A. J. Chadwick, "Autogenic erosional surfaces in fluvio-deltaic stratigraphy from floods, avulsions, and backwater hydrodynamics," *Journal of Sedimentary Research*, vol. 89, no. 8, pp. 815–832, 2019.
- [14] M. Nones, "Dealing with sediment transport in flood risk management," *Acta Geophysica*, vol. 67, no. 2, pp. 677–685, 2019.



- [15] J. M. Johnson, D. Munasinghe, D. Eyelade et al., “An integrated evaluation of the national water model (NWM)–Height above nearest drainage (HAND) flood mapping methodology,” *Natural Hazards and Earth System Sciences*, vol. 19, no. 11, pp. 2405–2420, 2019.
- [16] M. Lippe, M. Bithell, N. Gotts et al., “Using agent-based modelling to simulate social-ecological systems across scales,” *GeoInformatica*, vol. 23, no. 2, pp. 269–298, 2019.
- [17] L. T. H. Binh, N. V. Umamahesh, and E. V. Rathnam, “High-resolution flood hazard mapping based on nonstationary frequency analysis: case study of Ho Chi Minh City, Vietnam,” *Hydrological Sciences Journal*, vol. 64, no. 3, pp. 318–335, 2019.
- [18] H. M. Lyu, S. L. Shen, A. Zhou et al., “Perspectives for flood risk assessment and management for mega-city metro system,” *Tunnelling and Underground Space Technology*, vol. 84, pp. 31–44, 2019.
- [19] X. K. Jacob, D. S. Bisht, C. Chatterjee et al., “Hydrodynamic modeling for flood hazard assessment in a data scarce region: a case study of bharathapuzha river basin,” *Environmental Modeling & Assessment*, vol. 25, no. 1, pp. 97–114, 2020.
- [20] R. Diedhiou, S. Sambou, S. Kane et al., “Calibration of HEC-RAS model for one dimensional steady flow analysis—a case of Senegal river estuary downstream diama dam,” *Open Journal of Modern Hydrology*, vol. 10, no. 3, pp. 45–64, 2020.
- [21] R. Hu, F. Fang, P. Salinas et al., “Numerical simulation of floods from multiple sources using an adaptive anisotropic unstructured mesh method,” *Advances in Water Resources*, vol. 123, pp. 173–188, 2019.

## Research Article

# Cluster Coordination between High-speed Rail Transportation Hub Construction and Regional Economy Based on Big Data

**Liang Zhao**  and **Yuanhua Jia**

*School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China*

Correspondence should be addressed to Liang Zhao; 18114059@bjtu.edu.cn

Received 21 October 2020; Revised 22 December 2020; Accepted 12 January 2021; Published 30 January 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Liang Zhao and Yuanhua Jia. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As people's lives get better and better, more and more people choose to travel and with that comes the demand for more transportation. For now, traditional transportation hubs can temporarily meet people's travel needs. If driven by big data concepts and methods, the various capabilities of high-speed rail transportation hubs will be sublimated, and the regional economy will be in line with the prosperity of this place. Proportionally, railway hubs are extremely attractive to the rapid growth of the regional economy. This paper takes the high-speed railway hub construction model under big data as the research object and verifies the reliability of the research model and the development of economic regions based on the high-speed railway data in recent years as reference parameters. This article selects the panel data of railway transportation and regional economy in China's provinces for 10 consecutive years from 2011 to 2020. Among them, seven indicators were selected for railway transportation: passenger volume, freight volume, passenger turnover, cargo turnover, number of railway employees, railway transportation industry fixed asset investment and construction scale, and per capita railway network density. In terms of regional economy, six indicators were selected: regional GDP, per capita GDP, per capita investment in fixed assets, per capita total retail sales of consumer goods, per capita investment in imports and exports, and the proportion of the added value of the tertiary industry in GDP. The experimental results prove that each sample is tested in pairs, the standard error level of the mean is 0.002, which is less than 0.05, and high-speed railway construction can finally achieve economic integration. By improving the development of high-speed railways, continuously shortening the distance between time and space, breaking regional trade barriers, and reducing the cost of commodity circulation, industrial interaction and coordinated development between different regions can be effectively promoted.

## 1. Introduction

Rapid economic growth has led to a large amount of transportation demand, and differences in travel and service requirements have gradually diversified transportation and established a complete transportation hub. This is consistent with the ideas and methods of big data. How to use big data ideas and technologies to effectively use this information to serve the construction, management, and operation of nodes has become an important issue for improving the service level of integrated transmission nodes.

The development of railway transportation abroad has existed for a long time, and the research on the relationship between railway transportation and regional economic

development is relatively mature. China's railway construction is relatively late, but, with the construction and operation of railway transportation in various parts of our country, the relationship between railway transportation and the coordinated development of regional social and economic development has attracted more and more attention. There is a strong interdependence between a region and its neighbors. Therefore, the economic development of a region depends not only on its own foundation and investment but also on the development history of other regions. Li et al. pointed out that participation in high-level education and R&D has a higher impact on economic welfare. Furthermore, R&D indicators can be used as an important driving factor for evaluating economic welfare

[1]. When Wang and Loo studied the travel time of passengers, they studied the impact of high-speed railway on the time and space distance in the city network and derived the relationship between travel time and regional accessibility [2]. Pascal and Souza obtained the construction and operation of the high-speed railway through the research of the French high-speed railway TGV, which can improve regional accessibility [3].

Domestic researchers later studied the impact of high-speed railway construction on the local economy, but there are few studies. However, with the rapid development of China's high-speed railway construction and operation, research in this field has gradually been enriched and certain results have been achieved. With the improvement of technology and the wide application of measurement methods, scholars began to use quantitative analysis tools to quantitatively study the coordination relationship between railway transportation and regional economy. Based on the status quo of China's railway transportation and economic development, foreign theories and methods are used to conduct relevant research and analysis in different regions. Ikeda et al. used the industrial agglomeration theory of new economic geography to analyze the differences in China's regional economic development. It has been found that, since joining the World Trade Organization, the Gini coefficient where the industry is located has been declining, and the changes in regional economic differences have slowed down [4]. Han et al. used the linear weighting model and the dispersion coefficient method to quantitatively study the dynamic coordinated development level of the regional system and evaluated the effective degree of coordination and development within the regional system and between subsystems [5]. Li et al. established a dynamic model through dynamics and analyzed the relationship between transportation and economy [6].

Based on the research results at home and abroad, this paper finds that the algorithm of each method is based on different research data. This paper constructs a high-speed railway regional economic development model and studies route planning and high-speed railway integration models for big data. This paper conducts cluster analysis on three types of data and analyzes that the overall dimensionality reduction based on four weighting algorithms can effectively remove the attributes of less important research objects.

## 2. Cluster Coordination between High-Speed Railway Transportation Hub Construction and Regional Economy Based on Big Data

### 2.1. High-Speed Railway Transportation Hub, Hub Informatization, and Big Data

*2.1.1. Informatization of High-Speed Railway Transportation Hub.* Renewing the high-speed railway hub is very important to the management of the hub. Information management can further highlight the benefits of various transportation, high-speed railway, effective and targeted solutions for node management problems, and improving the service level of the transmission system. In terms of

coordinating transportation methods, handling emergency situations, and managing transportation needs, the management of high-speed railway hubs faces many management problems [7, 8]. To create a high-speed railway hub, we must propose targeted solutions to these management problems. Currently, a large number of nodes have created node information management platforms; these platforms have made significant contributions to coordinate link management, emergency, and rapid response and guide the flow of passengers. After determining the two basic conditions, use the reachability calculation model for calculation.

*2.1.2. High-Speed Railway Transportation Hub and Big Data.* From the current general point of view, big data has the characteristics of large quantity, variety, and speed, which represents a huge data scale, complex and diverse data forms, and the possibility of accurate and fast processing [9, 10]. Due to specific traffic functions and design locations, high-speed railway hubs have the characteristic of generating a large amount of different and complex data at high speed. For the city's large passenger transport hub, hundreds of thousands of passengers enter the hub every day. Faced with a large number of passenger transportations and logistics, high-speed railway operators have to deal with a large amount of complex information from video surveillance systems [11]. Big data indicates higher requirements for information: more information, more complete types of information, and more effective processing methods.

*2.2. Regional Economic Coordination.* First, construct the database according to the national economic census report specifications and prepare the network data report; then each census unit fills in the network report according to the nature of the unit and reports it directly, and each basic statistics bureau or department reviews the reported data and then reports it level by level. Analyze and process the final reported data, dig out the laws of economic development, discover economic development problems, and use modern information technology to make scientific predictions to form an analysis report [12, 13]. The economic census data processing process mainly includes four steps: report preparation, data collection, summary analysis, and data release.

*2.2.1. Improve Regional Accessibility.* Regional accessibility is reflected in the diversity of ways to reach a city. The improvement of traffic conditions has shortened the space-time distance between cities, and the frequency and spatial distance of people's travel have increased a lot compared to before. The convenient transportation environment improves the accessibility of the region and increases the attractiveness of the region. More enterprises and productivity are attracted to inject new vitality into the development of the region [14, 15]. The station has improved the city's traffic conditions and increased the radius of the city's activities and exchanges. The increase in speed and the reduction in

travel distance have resulted in cross-regional work and schooling.

**2.2.2. The Impact of High-Speed Railway on Regional Integration.** The integrated big data center collects a large amount of data from multiple sources in the transportation industry, such as urban roads, railway transportation, land public transportation, leasing, railway passenger and freight transportation, and freight Hadoop to create large data processing groups. Large amounts of traffic data are stored and processed [16, 17]. The integrated big data center is mainly divided into four databases: basic database, business database, exchange database, and theme database, used to store structured and unstructured data and, on this basis, carry out statistical analysis, web analysis, and data mining. The big data of the completed transmission node is waiting to be processed.

**2.3. Method Based on Clustering Feature Summary.** For the original data set that is too large to be read into memory at one time, you can first identify and summarize some of the data objects, reduce their memory usage, and complete large-scale data cluster analysis. Classical methods can be used, such as hierarchical method of balanced iterative induction clustering algorithm [18, 19]. The BIRCH clustering algorithm uses the CF tree to summarize information. CF is the main component of the node of the tree and the core of BIRCH. All the information of the cluster in the CF includes all the information needed for clustering. If you want to merge two clusters, you only need the sum of their corresponding CFs.

$$\begin{aligned}\bar{X}_0 &= \frac{\sum_{i=1}^N \bar{x}_i}{N}, \\ R &= \left( \frac{\sum_{i=1}^N (\bar{x}_i - \bar{x}_0)^2}{N} \right)^{1/2}, \\ D &= \left( \frac{\sum_{i=1}^N \sum_{j=1}^N (\bar{x}_i - \bar{x}_j)^2}{N(N-1)} \right)^{1/2}.\end{aligned}\quad (1)$$

The above information can be used to represent the information of the internal members of a cluster. The distance between clusters is used to express the degree of separation between clusters [20]. BIRCH defines the following 5 types of distance formulas between clusters. These distance formulas describe the degree of separation between clusters from different angles.

$$\begin{aligned}D_0 &= \left[ (\bar{x}_i - \bar{x}_j)^2 \right]^{1/2}, \\ D_1 &= |\bar{X}_{0_1} - \bar{X}_{0_2}| = \sum_{i=1}^N |\bar{X}_{0_1}^{(i)} - \bar{X}_{0_2}^{(i)}|, \\ D_2 &= \left( \frac{\sum_{i=1}^{N_1} \sum_{j=N_1+1}^{N_1+N_2} (\bar{x}_i - \bar{x}_j)^2}{N_1 N_2} \right)^{1/2}, \\ D_3 &= \left( \frac{\sum_{i=1}^{N_1+N_2} \sum_{j=1}^{N_1+N_2} (\bar{x}_i - \bar{x}_j)^2}{(N_1 + N_2)(N_1 + N_2 - 1)} \right)^{1/2}, \\ D_4 &= \sum_{j=N_1+1}^{N_1+N_2} \left( \bar{x}_j - \sum_{i=N_1+1}^{N_1+N_2} \frac{\bar{x}_i}{N_2} \right)^2.\end{aligned}\quad (2)$$

Because CF completely retains some of the key information required for the above calculations, BIRCH can use CF to complete large-scale data clustering. For example, CF can calculate the cluster radius  $R$ , where  $R = (((SS - \overline{LSLS}^T)/N)/N)^{1/2}$ . Similarly, the diameter  $D$  and the distance between clusters  $D_0 - D_4$  can be calculated using CF. It can be calculated by the six following formulas:

$$\begin{aligned}D &= \left( \frac{2N(SS) - 2\overline{LSLS}^T}{N(N-1)} \right)^{1/2}, \\ D_0 &= \left[ \left( \frac{\overline{LS}_1}{N_1} - \frac{\overline{LS}_2}{N_2} \right)^2 \right]^{1/2}, \\ D_1 &= \left| \sum_{i=1}^d \left( \frac{1}{N} \overline{LS}_1^{(i)} - \frac{1}{N_2} \overline{LS}_2^{(i)} \right) \right|, \\ D_2 &= \left( \frac{N_2 SS_1 - 2\overline{LS}_1 \overline{LS}_2^T + N_1 SS_2}{N_1 N_2} \right)^{1/2}, \\ D_3 &= \left( \frac{2(N_1 + N_2)(SS_1 + SS_2) - 2(\overline{LS}_1 \overline{LS}_1^T + \overline{LS}_2 \overline{LS}_2^T)}{(N_1 + N_2)(N_1 + N_2 - 1)} \right)^{1/2}, \\ D_4 &= \frac{\overline{LS}_1 \overline{LS}_1^T}{N_1} + \frac{\overline{LS}_2 \overline{LS}_2^T}{N_2} - \frac{\overline{LS}_1 \overline{LS}_1^T + \overline{LS}_2 \overline{LS}_2^T}{N_1 + N_2}.\end{aligned}\quad (3)$$

Dice similarity coefficient calculation formula is as follows:

$$\text{DICE} = \frac{2|\text{SEG} \cap \text{GT}|}{|\text{SEG}| + |\text{GT}|},$$

$$y = \alpha W y + \beta_1 X - W \beta_2 X + \varepsilon, \quad (4)$$

$$\ln g_{it} = \alpha_0 + \alpha_1 du * dt + \sum_{i=1}^N b_j Xu + \varepsilon_u.$$

SEN coefficient calculation formula is as follows:

$$\text{SEN} = \frac{|\text{SEG} \cap \text{GT}|}{\text{GT}}, \quad (5)$$

$$\theta(p, q) = \arctan\left(\frac{L(p, q+1) - L(p, q-1)}{L(p+1, q) - L(p-1, q)}\right), \quad (6)$$

$$\psi = \sum_{x=1}^{\theta} Vx = \sum_{x=1}^{\theta} \left( \frac{Wx}{\sum_{x=1}^x Wx} Sx \right). \quad (7)$$

In the process of clustering, if you want to cluster efficiently, you need to determine what is important data and what is secondary data; important data is very important to the role of the cluster, and secondary data can be processed secondarily [21]. Keep the primary data, summarize the secondary data, and save their clustering information with appropriate statistics.

## 2.4. Conception of Hub Informatization Based on Big Data

**2.4.1. Discover More Needs and Patterns from the Data.** The high-tech data analysis technology of this node will enable managers to proactively investigate traffic demand without waiting for passengers to make requests or even solve the supply-demand relationship when problems arise. Due to its content, the integrated transportation node has different operation modes and characteristics from general transportation methods and transportation stations. Their completeness and complexity make it difficult to penetrate many node information processing methods. With the help of big data ideas and technologies, it can not only process and publish diverse, comprehensive, and complex data but also provide valuable resources. For administrators and service providers, it is undoubtedly beneficial to discover the tacit traffic demand from these information-rich resources and discover the changing mode of the mobile center [22, 23]. Mined needs and patterns can also provide targeted and personalized traffic services for functional groups and scenarios. With the help of big data, it will be possible to meet the needs of traffic service users and improve the experience of using traffic services.

**2.4.2. More Accurate Warning and Faster Response.** Emergency management is the most difficult part of managing an integrated transportation hub. In emergency situations, the quick response level can usually tell you the advantages and disadvantages of node management. The node information management method can help managers to be calmer and effectively deal with emergencies and can

help managers to develop better evacuation plans and more effective passenger flow guidance. With the help of big data, the emergency response level of the center will be significantly improved. This is reflected not only in the effectiveness after emergence but also in effective early warning. Standard mode analysis of multiple data conditions will help establish an effective and accurate event warning mechanism. By analyzing a large amount of emergency data at historical nodes [24, 25], the emergency mode can be obtained, and then the phenomenon warning can be compared with the operation management, or the display mechanism can be further analyzed by the following method: mode to achieve event warning and quick response effect. With the help of big data processing methods, real-time monitoring can also quickly make faster and more accurate judgements about the state of the center.

**2.4.3. Basis of Complex Coordinated Linkage Technology.** Efficient data processing mechanism will lay the foundation for complex adjustment and link technology. Large-scale data analysis and processing methods will enable integrated transmission nodes to quickly process this data while receiving large amounts of data. The complexity of a node determines the complexity of its data structure. The data used by the management and transmission service provider must also be processed before it can be released and exchanged with the connection provider. Under normal circumstances, only simple data such as train arrivals and messages can be transmitted in real time. Slightly more complex data interactions require at least one day, even a week, or a month to be transmitted to the other party [26, 27]. At present, due to lack of timeliness, it is difficult to provide help for the current services of the node, and most of them are used to summarize experience. The mechanism for quickly processing large amounts of data can quickly improve and transmit effective information and even interact to achieve the goal of high-quality services. Within the scope of application of big data technology, better coordination and link management will be carried out.

## 2.5. Influence of Railway on Regional Economy

**2.5.1. The Influence of Railway on the Formation and Development of China's Regional Economic Belt.** China's railway development process reflects the course of China's economic development from the side. Judging from its early military and political importance or its successor economic importance, railway transportation has always played an important role and has made indelible achievements in China's economic development. It is obvious to promote the development of railway in the development of regional economy. The development of regional economy has always been closely connected with transportation, which requires strong transportation guarantee. As the main force of transportation, railway transportation has an irreplaceable status. Since the reform and opening up, railway transportation has undergone many major developments and accelerations to realize the effective circulation of human

resources, logistics, and other market elements. The construction and continuous improvement of the railway network have promoted the development of China's southeast coastal areas and linked the supply of raw materials with the central and western regions. It can be understood that railway transportation is a necessary and sufficient condition for the formation of a regional economy, the first construction of railway transportation, and the subsequent formation of the regional economy. Looking at the main railway lines in China, it is not difficult to see that they connect China's central cities to develop urban economic and commercial circles and guide urban development. For example, as transportation hubs, central cities such as Shanghai and Beijing have played a role in economic radiation and have become continuous trade corridors.

*2.5.2. Links between Railways and Major Economic Regions.* Railway transportation has played an important role in China's economic development. Significant changes have taken place in the industrial layout of areas along the railway, and railways have become an important part of the industrial layout. After the reform and opening up, we have observed the process of China's economic development, and it is not difficult to see that important industries are concentrated in areas along the main railway lines. The development of large railways has paved the way for the establishment of economic zones, and the layout of the railway network has also contributed to the economy and has had a significant impact on development. So far, the strategic planning for the distribution of productivity along the main railways in Northeast, North China, Southeast, Northwest, and Southwest has been completed, and the development of railways has become the political guide for economic development strategies. Facts fully show that the gradual maturity of urban settlements will stimulate interregional traffic demand, and the increasing traffic demand will in turn promote the connections and exchanges between urban settlements and will also develop regional economies.

*2.5.3. Influence of the Railway on the Economic Structure Adjustment of the Areas along the Route.* Passenger and cargo exchange in railway transportation is an important factor in economic growth. The construction and sustainable development of the railway will inevitably have an impact on the economic structure of the regions along the route, and the subsequent economic growth along the route will also expand the total regional economy. The direction of the railway also reflects the layout of the resource flow. The resource line can promote the formation of the regional economy by promoting the development and upgrading of industry and providing appropriate materials and labor.

*2.5.4. Influence of Railway on the Economic Development of Underdeveloped Areas.* Railway traffic has a unique role in promoting the economy of underdeveloped areas along the line. Take a typical case in the economic development of the United States to illustrate that the development of natural

resources and economic development in the western region of the United States benefited from the large-scale railroad construction in the United States in modern times, especially the rapid development of heavy industrial industries such as steel and machinery related to the railway industry. Our country's economic development has always been uneven. Since the reform and opening up, the economic gap between regions has gradually widened. Economically active areas have more human resources, modernization and economic development levels are higher, but the geographical space is smaller, and more regional space is needed; meanwhile those economically backward areas have abundant natural resources and vast geographical space but scarce economic growth point. Transportation is the main channel and important link for communication between these two types of regions, which organically connects the economically developed and economically backward regions, carries out the flow and transfer of economic factors between them, and effectively transfers capital, technology, and talents.

### **3. Coordination between the Construction of High-Speed Railway Transportation Hub and Regional Economy**

*3.1. Demand Analysis.* Intelligent data analysis is based on data collection, statistics, and aggregation and uses intelligent analysis models to mine and discover census data relationships, economic development issues, economic level regional division, economic index level status, and national economy based on economic indicators development layout countermeasures and suggestions. Intelligent data analysis needs to meet the needs of effective statistics and clustering of data reports that cannot be counted by existing database technology, the need to assist database technology in order to find problems more comprehensively, so as to have a more comprehensive grasp of the overall situation, and the need to plan for the next step of economic development basic analysis work and forecast suggestions.

*3.2. Database Storage System.* The database storage system is the foundation of the entire system and a prerequisite for intelligent analysis of census data. This database storage system mainly includes functions such as user management, data entry and import and export, data query and summary, system navigation assistance, and system security management.

*3.3. Test Subject.* The intelligent data analysis system mainly completes the intelligent analysis of economic census data, and its goal is to use intelligent analysis methods and data mining related technologies to discover unknown laws in economic census data. The system tried four methods of intelligent analysis of the economic census data, fuzzy clustering, MMD algorithm, K-means algorithm, and FCM algorithm, and analyzed three types of data: economic growth, financial status, employment by industry and wages. A cluster analysis was carried out and a good conclusion was obtained. At the same time, data from railway traffic and

regional economic data tables are selected for 10 consecutive years from 2011 to 2020. Among them, the railway traffic data comes from the “National Railway Statistics Abstract 2011–2020,” the Ministry of Transport, and regional economic data from China’s 2011–2020 statistical year, provincial and municipal statistical yearbooks, and the National Bureau of Statistics. It provides practical value that can be reported at the decision-making level and can refer to information that cannot be obtained by traditional database analysis methods and statistical methods.

**3.4. Experimental Method.** This paper selects the panel data of China’s provincial railway traffic and regional economy for 10 consecutive years from 2011 to 2020. Among them, railway transportation selects seven indicators: passenger volume, freight volume, passenger turnover, cargo turnover, number of railway employees, railway transportation industry fixed asset investment construction scale, and per capita railway network density. In terms of regional economy, six indicators are selected: regional GDP, per capita GDP, per capita fixed asset investment, per capita retail sales of consumer goods, per capita import and export investment, and the proportion of tertiary industry added value in GDP.

**3.5. Data Processing.** The index selected in this article is the annual statistical data of books, which is the index data of practical value. Therefore, the trapezoidal fuzzy table distribution function is usually used to calculate the participation of each indicator. The specific formula is as follows: partially large trapezoidal distribution is as follows:

$$r(x) = \frac{x-c}{d-c}, \quad c < x < d. \quad (8)$$

Partially small trapezoidal distribution is as follows:

$$r(x) = \frac{b-x}{x-a}, \quad a < x < b. \quad (9)$$

For the ideal score value, it can usually be set as the middle value of each interval. This paper divides the evaluation grades into five grades, and the index data after normalization processing all fall into the [0-1] interval.

There are many ways of data standard processing, but different data standardization methods will have a certain impact on the evaluation results of the system. For the positive index standardization method,

$$y_{ij} = \frac{x_{ij} - \min(x_{ij})}{\max(x_{ij}) - \min(x_{ij})}. \quad (10)$$

For negative index standardization methods,

$$y_{ij} = \frac{\min(x_{ij}) - x_{ij}}{\max(x_{ij}) - \min(x_{ij})}. \quad (11)$$

After standardizing the data, using the principal component analysis of nonlinear logarithmic centering, the

processing steps of logarithmic transformation and row vector centering are

$$z_{ij} = \ln y_{ij} - \sum_{i=1}^m \frac{\ln y_{ij}}{m}. \quad (12)$$

**3.6. Statistical Methods.** SPSS 23.0 software was used for data processing, and the count data was expressed as a percentage (%),  $k$  is the amount of data in this experiment,  $\sigma^2$  is the variance of all survey results, and  $P < 0.05$  indicates that the difference is statistically significant. The formula for calculating reliability is shown in equation (7).

$$a = \frac{k}{k-1} \left( 1 - \frac{\sum \sigma_i^2}{\sigma^2} \right). \quad (13)$$

## 4. Coordination between the Construction of High-Speed Railway Transportation Hub and Regional Economy

**4.1. Evaluation Index System Based on Index Reliability Testing.** Reliability refers to the stability and reliability of the questionnaire. This article adopts the  $\alpha$  coefficient method created by L. J. Cronbach. The  $\alpha$  coefficient can be obtained by Reliability Analysis in SPSS software. It is generally believed that the  $\alpha$  coefficient above 0.8 indicates that the effect of index setting is very good, and that above 0.7 is also acceptable.

This experiment can be used to perform cluster analysis on three types of data: economic growth, financial status, employment status by region and wages ( $\alpha > 0.7$ ). As shown in Table 1, the economic growth under the fuzzy clustering model is 0.8427. There are no absolute pros and cons between analyses, not to say that the more complex the model, the better the prediction performance. Within an acceptable range, the preconditions of the experiment are met, which provides a basis for subsequent experimental analysis.

**4.2. Three Indicators by Different Intelligent Analysis Methods.** As shown in Figure 1 and Table 2, economic attributes are mainly divided into three subitems: economic growth, financial status, and employment of personnel by industry and wages by region. There are 79 dimensions in total. This time, 4 algorithms are used for overall dimension reduction, which overcome the shortcomings of inaccurate clustering results of different densities by general algorithms.

It can be seen from Figure 2 and Table 3 that the overall dimensionality reduction based on the four weighted algorithms can effectively remove the attributes of the research object with less significance. It is similar to the principal component analysis and the multiattribute projection pursuit method, but the effect is better after clustering.

As shown in Table 4, the result is more accurate. Here, choose the intelligent analysis of four methods: fuzzy clustering, MMD algorithm, K-means algorithm, and FCM algorithm, and perform cluster analysis on the three types of

TABLE 1: Summary table of reliability test results.

| Intelligent analysis | Type of data  | Alpha coefficient ( $\alpha$ ) |
|----------------------|---|--------------------------------|
| Fuzzy clustering     | Economic growth   | 0.8427                         |
| MMD algorithm        | Financial situation   | 0.7652                         |
| K-means algorithm    | Personnel employment and personnel salary by industry by region | 0.7372                         |
| FCM algorithm        |   | 0.7592                         |

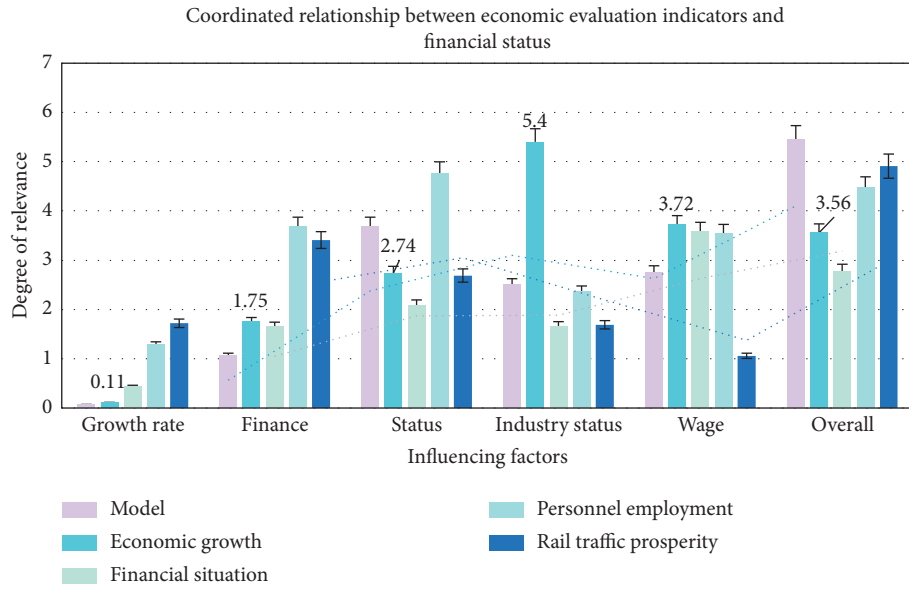


FIGURE 1: Coordinated relationship between economic evaluation indicators and financial status.

TABLE 2: The construction and analysis of the high-speed railway economic hub.

| Item            | Model | Economic growth | Financial situation | Personnel employment | Rail traffic prosperity |
|-----------------|-------|-----------------|---------------------|----------------------|-------------------------|
| Growth rate     | 0.07  | 0.11            | 0.44                | 1.28                 | 1.72                    |
| Finance         | 1.06  | 1.75            | 1.66                | 3.69                 | 3.41                    |
| Status          | 3.69  | 2.74            | 2.09                | 4.76                 | 2.69                    |
| Industry status | 2.5   | 5.4             | 1.67                | 2.36                 | 1.69                    |
| Wage            | 2.75  | 3.72            | 3.59                | 3.55                 | 1.06                    |
| Overall         | 5.46  | 3.56            | 2.78                | 4.47                 | 4.91                    |

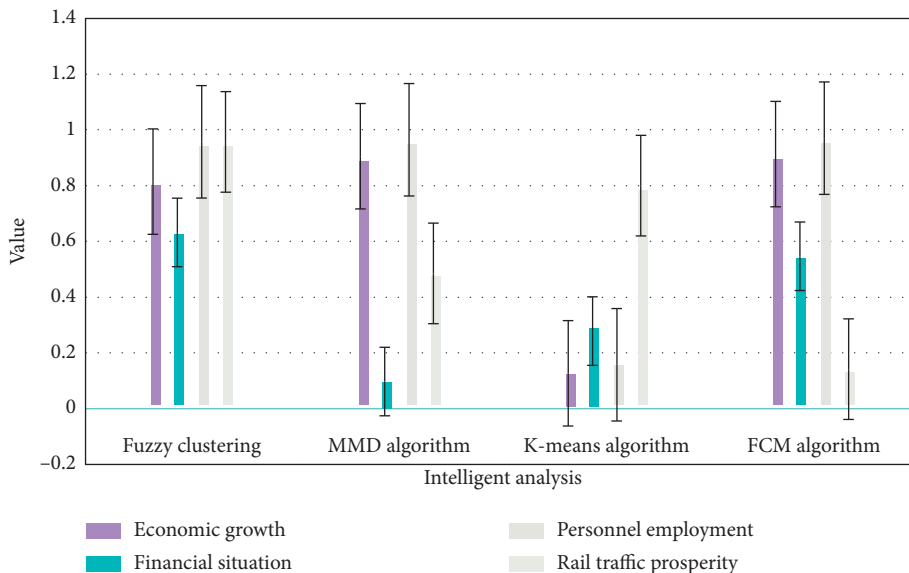


FIGURE 2: Analysis diagram of three indicators by different intelligent analysis methods.



TABLE 3: Three indicators by different intelligent analysis methods.

| Model            | Economic growth | Financial situation | Personnel employment | Rail traffic prosperity |
|------------------|-----------------|---------------------|----------------------|-------------------------|
| Fuzzy clustering | 0.8147          | 0.6324              | 0.9575               | 0.9572                  |
| MMD algorithm    | 0.9058          | 0.0975              | 0.9649               | 0.4854                  |
| K-means          | 0.127           | 0.2785              | 0.1576               | 0.8003                  |
| FCM algorithm    | 0.9134          | 0.5469              | 0.9706               | 0.1419                  |

TABLE 4: The growth level of the regional economy and the development of the railway hub.

| Model            | Economic growth | Financial situation | Personnel employment | Rail traffic prosperity |
|------------------|-----------------|---------------------|----------------------|-------------------------|
| Fuzzy clustering | 0.8147          | 0.6324              | 0.9575               | 0.9572                  |
| MMD algorithm    | 0.9058          | 0.0975              | 0.9649               | 0.4854                  |
| K-means          | 0.127           | 0.2785              | 0.1576               | 0.8003                  |
| FCM algorithm    | 0.9134          | 0.5469              | 0.9706               | 0.1419                  |

data of economic growth, financial status, and regional employment and wages. It proves that the growth level of the regional economy and the development of the railway hub have a mutually promoting effect.

As shown in Figure 3 and Table 5, the scope of the overall planning of innovation resources at the city level has been expanded to form a hub innovation circle within 10 minutes of a comprehensive transportation hub site. The change in the spatial form of the innovation block at the block level has formed a TOD innovation block within an 800 m walk-in radius of a high-intensity development area. The vertical concentration of innovative functions at the architectural level is a comprehensive innovation core that uses the complex as a carrier to gather innovative elements.

#### 4.3. Evaluation Model of System Coupling Coordination and Analysis of Results

**4.3.1. Judgement Analysis of the Correlation of Evaluation Indicators.** As shown in Table 6 and Figure 4, before further analysis of the data, first analyze the correlation of evaluation indicators and the partial correlation between variables. The correlation coefficient matrix and the corresponding statistical value are obtained through 4 algorithms. Judging whether the evaluation index is suitable for principal component analysis according to the value, KMO greater than 0.9 indicates very suitable; KMO less than or equal to 0.9 and greater than 0.8 indicates appropriate; KMO less than or equal to 0.8 and greater than 0.7 indicates appropriate; KMO less than or equal to 0.7 and greater than 0.6 indicates barely appropriate; KMO less than or equal to 0.6 and greater than 0.5 means inappropriate; KMO less than 0.5 means inappropriate.

As shown in Figure 5, the existence of high-speed railway and air service time windows will have a certain impact on the decision-making of multimodal transportation routes for fresh products. When the total cost is the goal, the “high-speed railway + highway” transportation combination

method can be selected in consideration of the node time window; in addition, compared with the optimal route without the node time window, the existence of the time window of the transportation mode at the node is to a certain extent. It increases the cost of time value, but it is more realistic and reasonable than the idealized scenario without time window in previous studies. Multimodal transport operators can choose appropriate routes and transportation methods according to the situation.

The result is shown in Figure 6 and Table 7; at the same time, each sample is tested on the selected samples, and the mean, standard deviation, standard error of the mean, and other data are obtained, as shown in Table 2.

It can be seen from Table 8 that, at the same time each sample is tested in pairs, the standard error level of the mean is 0.04, which is greater than 0.05, indicating that the correlation matrix of the evaluation indicators of the railway transportation system and the regional economic system is suitable for principal component analysis.

It can be seen from Figure 7 and Table 9 that the construction of high-speed railways is an important means to accelerate regional economic development. High-speed railways can not only adjust the industrial structure and improve the market structure but also deepen interregional connections, increase the employment rate, and inject the regional economy vitality.

#### 4.3.2. Railway Traffic Index Data Based on MMD Algorithm.

As shown in Figure 8 and Table 10, in terms of railway transportation, we select passenger volume, freight volume, passenger turnover, cargo turnover, railway employment, railway transportation industry fixed asset investment and construction scale, and per capita railway network density. These seven indicators mainly adjust the maximum number of iterations of the model, the learning rate, and the maximum depth of the tree. Each iteration will produce a weak learner. If the number of weak learners is too small, it is easy to underfit, and if there are too many, it is easy to overfit.

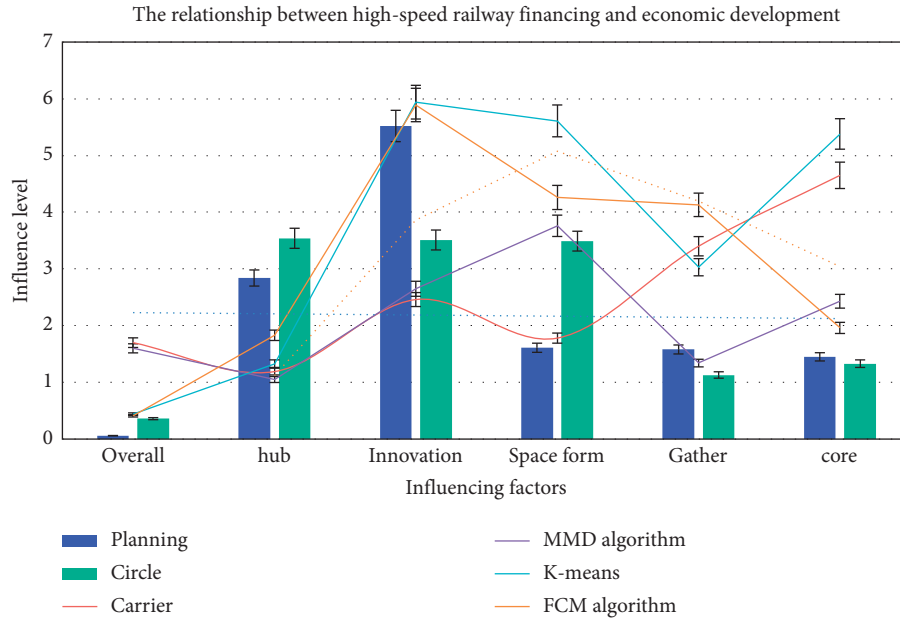


FIGURE 3: The relationship between high-speed railway financing and economic development.

TABLE 5: The overall planning of innovation resources.

| Item       | Planning | Circle | Carrier | MMD algorithm | K-means | FCM algorithm |
|------------|----------|--------|---------|---------------|---------|---------------|
| Overall    | 0.06     | 0.36   | 1.7     | 1.6           | 0.44    | 0.41          |
| Hub        | 2.84     | 3.54   | 1.19    | 1.05          | 1.33    | 1.83          |
| Innovation | 5.52     | 3.51   | 2.46    | 2.65          | 5.94    | 5.89          |
| Space form | 1.61     | 3.49   | 1.78    | 3.76          | 5.61    | 4.26          |
| Gather     | 1.58     | 1.13   | 3.4     | 1.34          | 3.03    | 4.13          |
| Core       | 1.45     | 1.33   | 4.65    | 2.43          | 5.38    | 1.96          |

TABLE 6: Relative index of economic evaluation of high-speed railway system.

| Item        | KMO  | Economic | Load | Space form | Gather | Core |
|-------------|------|----------|------|------------|--------|------|
| Safety      | 1.3  | 1.05     | 1.21 | 1.57       | 1.51   | 0.67 |
| Continued   | 1.92 | 1.66     | 3.2  | 2.89       | 1.97   | 3.81 |
| Development | 2.17 | 5.68     | 2.61 | 5.77       | 3.47   | 3.33 |
| Length      | 1.39 | 5.65     | 3.24 | 5.89       | 5.4    | 3.47 |
| Core        | 2.31 | 1.31     | 4.45 | 1.25       | 4.02   | 4.91 |

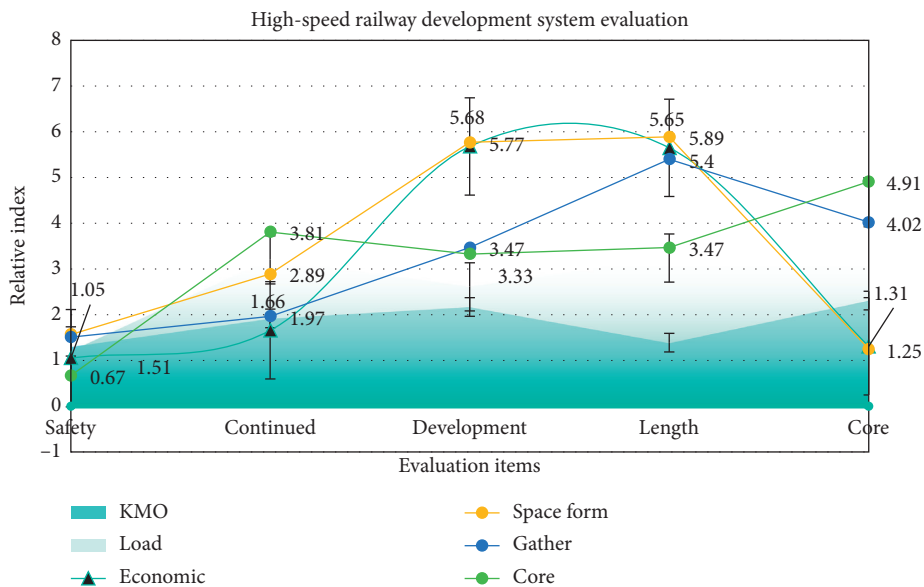


FIGURE 4: High-speed railway development system evaluation.

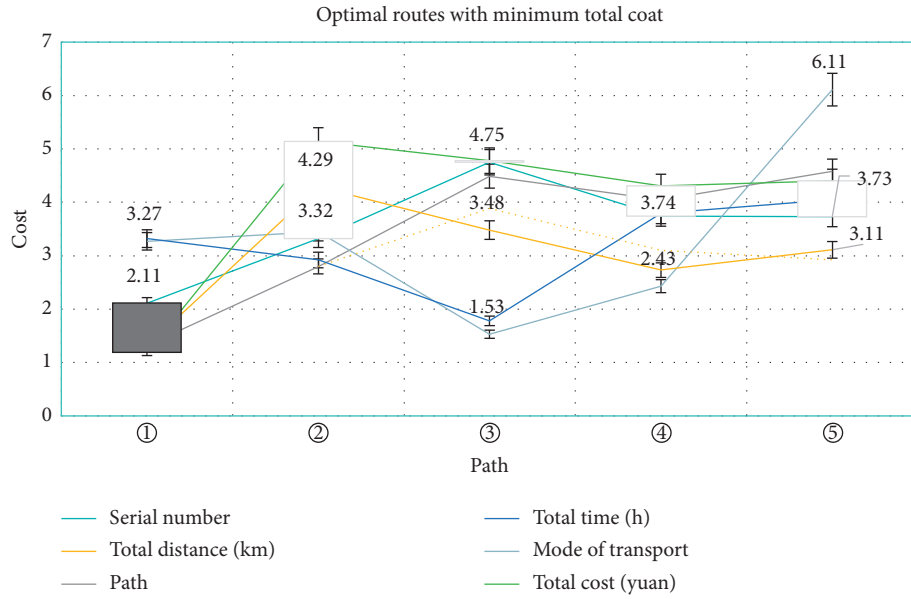


FIGURE 5: Optimal routes with minimum total cost (not existing time windows).

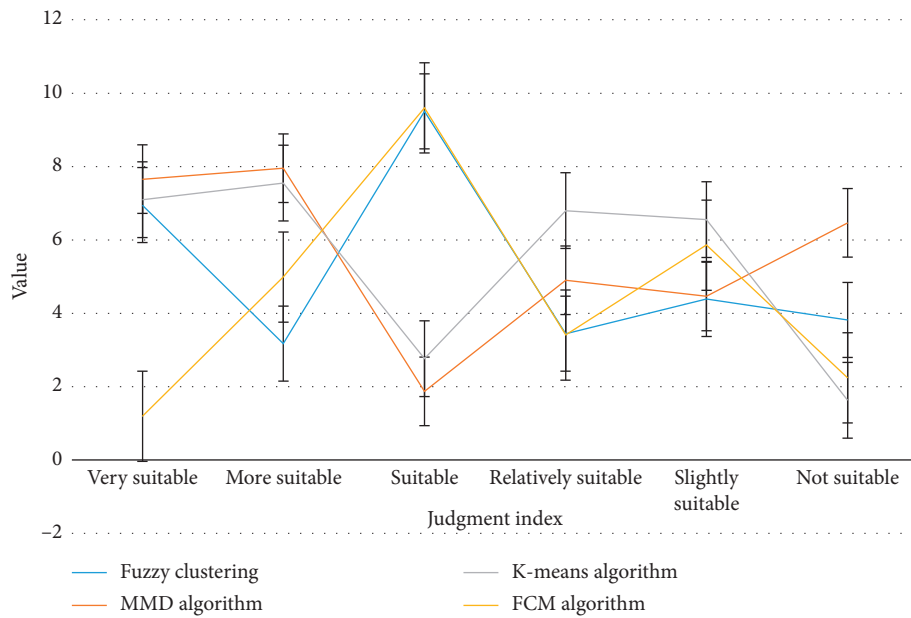


FIGURE 6: Judgement analysis diagram of evaluation index correlation.

TABLE 7: Optimal routes with minimum total cost.

| Num. | Serial number | Path | Mode of transport | Total distance (km) | Total time (h) | Total cost (yuan) |
|------|---------------|------|-------------------|---------------------|----------------|-------------------|
| ①    | 2.11          | 1.32 | 3.27              | 1.31                | 3.32           | 1.19              |
| ②    | 3.32          | 2.8  | 3.45              | 4.29                | 2.92           | 5.14              |
| ③    | 4.75          | 4.49 | 1.53              | 3.48                | 1.78           | 4.78              |
| ④    | 3.74          | 4.05 | 2.43              | 2.73                | 3.79           | 4.31              |
| ⑤    | 3.73          | 4.58 | 6.11              | 3.11                | 4.05           | 4.4               |

TABLE 8: Judgement table for the relevance of evaluation indicators.

|   | Mean  | Standard deviation | Standard error of the mean |
|---|-------|--------------------|----------------------------|
| Economic growth   | -1.42 | 8.347              | 1.624                      |
| Financial situation   | -2.64 | 7.324              | 1.628                      |
| Personnel employment and personnel salary by industry by region | 0.67  | 7.685              | 1.623                      |

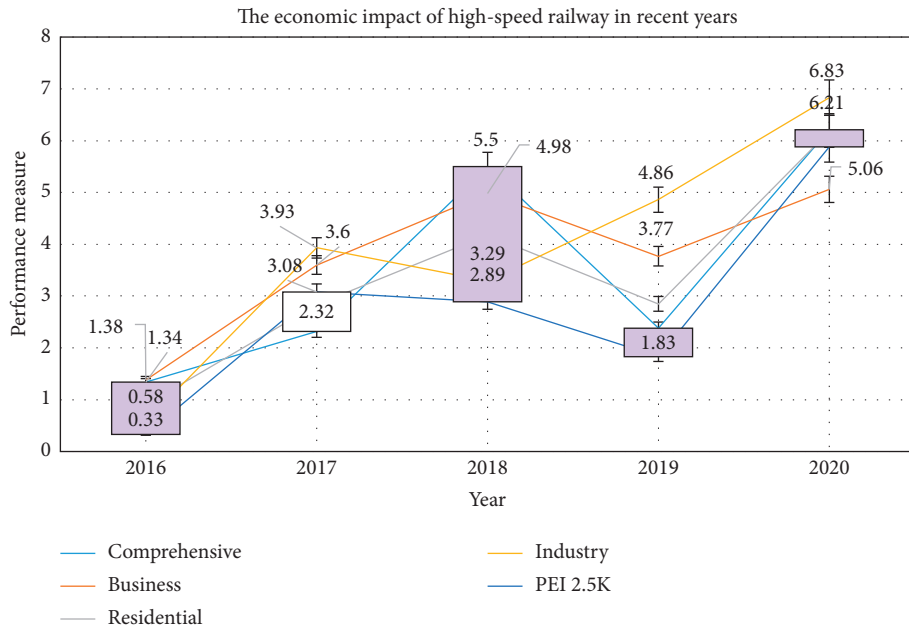


FIGURE 7: The economic impact of high-speed railway in recent years.

TABLE 9: Number of trials or attempts at learning.

| Year | Comprehensive | Business | Residential | Industry | PEI 2.5 K |
|------|---------------|----------|-------------|----------|-----------|
| 2016 | 1.34          | 1.38     | 0.9         | 0.58     | 0.33      |
| 2017 | 2.32          | 3.6      | 2.84        | 3.93     | 3.08      |
| 2018 | 5.5           | 4.98     | 4.18        | 3.29     | 2.89      |
| 2019 | 2.38          | 3.77     | 2.85        | 4.86     | 1.83      |
| 2020 | 6.21          | 5.06     | 6.18        | 6.83     | 5.88      |

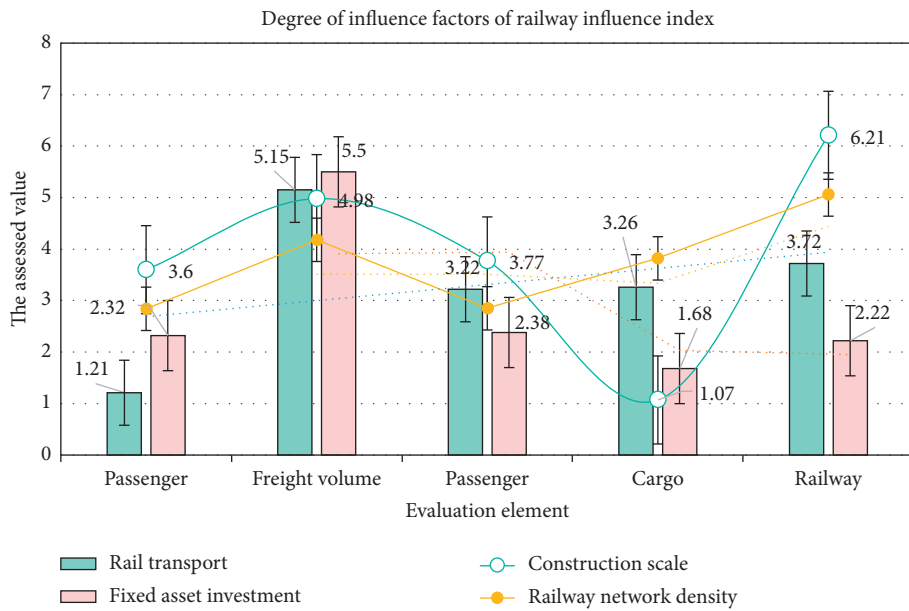


FIGURE 8: Degree of influence factors of railway influence index.

TABLE 10: Influencing factor data for railway impact indicators.

| Item           | Rail transport | Fixed asset investment | Construction scale | Railway network density |
|----------------|----------------|------------------------|--------------------|-------------------------|
| Passenger      | 1.21           | 2.32                   | 3.6                | 2.84                    |
| Freight volume | 5.15           | 5.5                    | 4.98               | 4.18                    |
| Passenger      | 3.22           | 2.38                   | 3.77               | 2.85                    |
| Cargo          | 3.26           | 1.68                   | 1.07               | 3.82                    |
| Railway        | 3.72           | 2.22                   | 6.21               | 5.06                    |

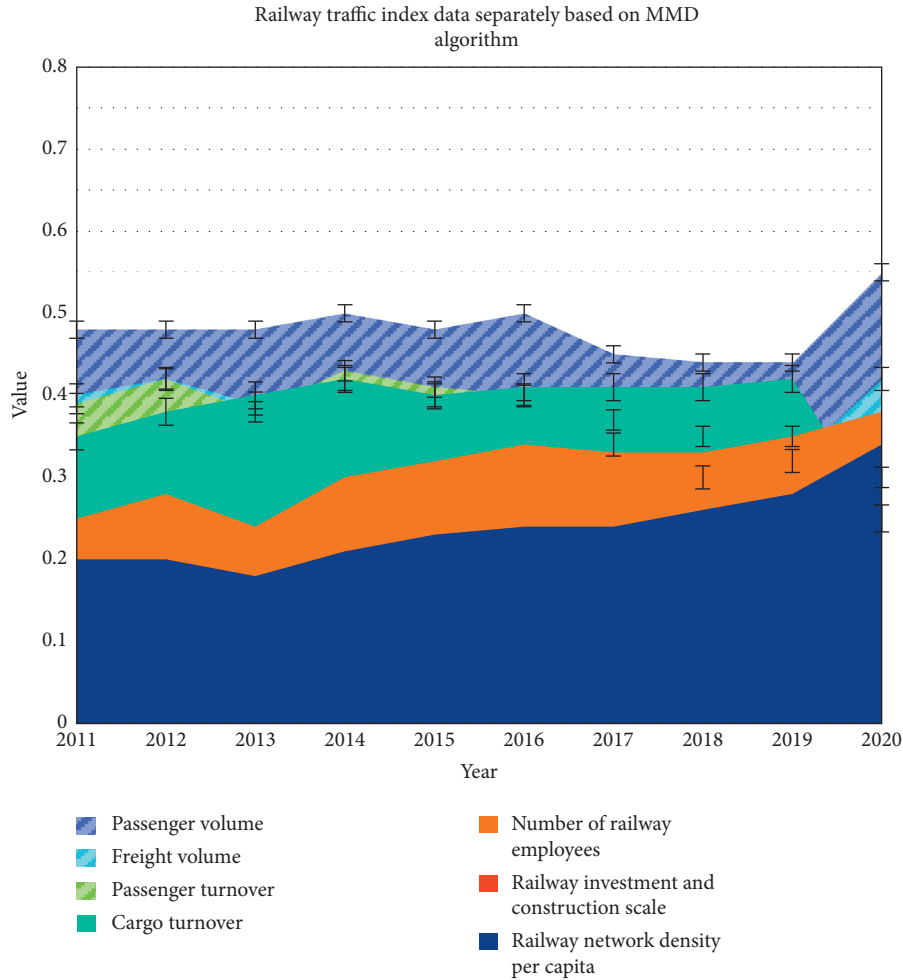


FIGURE 9: Research and analysis diagram of railway traffic index data separately based on MMD algorithm.

TABLE 11: Analysis table of the result analysis of the risk assessment system of Yebes algorithm.

|   | Mean  | Standard deviation | Standard error of the mean |
|---|-------|--------------------|----------------------------|
| Economic growth   | 2.32  | 9.462              | 1.526                      |
| Financial situation   | 1.94  | 8.347              | 1.524                      |
| Personnel employment and personnel salary by industry by region | -1.46 | 9.432              | 1.529                      |

TABLE 12: High-speed railway in recent years has prompted data.

| Year | Interpersonal communication | Advantage | Weak | Opportunity | Threat |
|------|-----------------------------|-----------|------|-------------|--------|
| 2016 | 0.63                        | 0.87      | 0.32 | 0.81        | 1.4    |
| 2017 | 1.39                        | 2.14      | 1.58 | 2.9         | 2.05   |
| 2018 | 2.29                        | 2.37      | 5.96 | 2.4         | 3.73   |
| 2019 | 4.16                        | 4.8       | 2.88 | 3.3         | 5.41   |
| 2020 | 2.28                        | 4.35      | 3.48 | 4.32        | 4.02   |

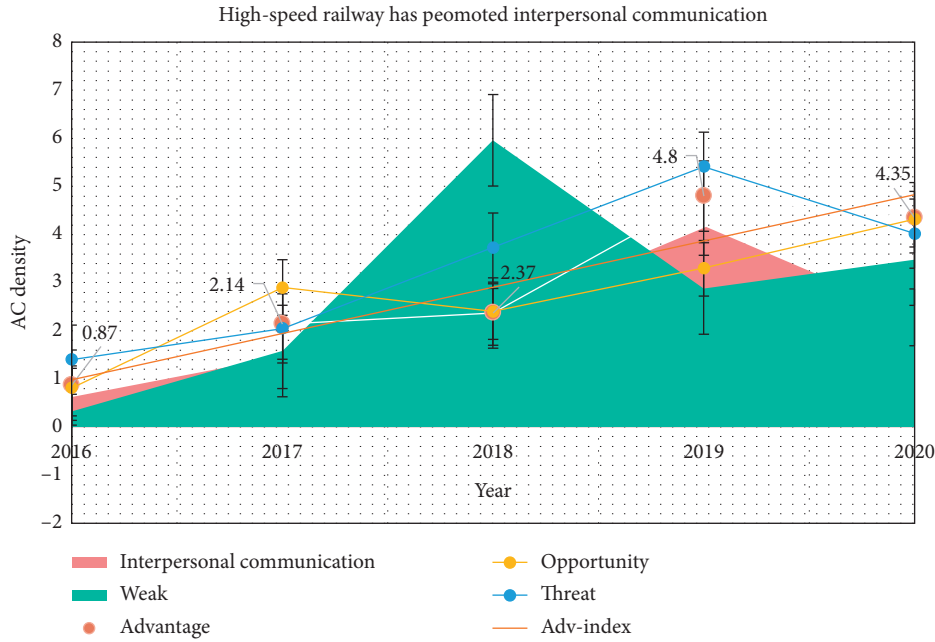


FIGURE 10: High-speed railway has promoted interpersonal communication.

The result is shown in Figure 9; at the same time, each sample is tested on the selected samples, and the mean, standard deviation, standard error of the mean, and other data are obtained, as shown in Table 11.

It can be seen from Table 12 that, at the same time each sample is tested in pairs, the standard error level of the mean is 0.002, which is less than 0.05, and high-speed railway construction can finally achieve economic integration. By improving the development of high-speed railways, continuously shortening time and space distances, breaking regional trade barriers, and reducing commodity circulation costs, industrial interaction and coordinated development among different regions can be effectively promoted.

As shown in Figure 10, relying on the high-efficiency and rapid advantages of high-speed railways drives the efficiency of personnel exchanges between different regions; through the release of existing line transportation capacity, the efficiency of cargo transportation is improved.

**4.3.3. Regional Economic Index Data Based on MMD Algorithm.** In terms of regional economy, we select six indicators of regional GDP, per capita GDP, per capita fixed asset investment, per capita retail sales of consumer goods, per capita import and export investment, and the proportion of tertiary industry added value in GDP and adjust the maximum number of iterations, learning rate, and maximum depth of the tree. Each iteration will produce a weak learner. As shown in Figure 11 and Table 13, if the number of weak learners is too small, it is easy to underfit, and if there

are too many, it is easy to overfit. Each sample is tested on the selected samples, and data such as the mean, standard deviation, and standard error of the mean are obtained, as shown in Table 14.

It can be seen from Figure 12 that, at the same time each sample is tested in pairs, the standard error level of the mean is 0.015, which is less than 0.05, indicating that, in the future development of high-speed railways, it is necessary to clarify and solve the problems since the opening of high-speed railways.

It can be seen from Figure 13 and Table 15 that, along the railway line, the effect of high-speed railways on the regional economy will be reflected to the greatest extent. The specific manifestation is the use of transfer stations and the radiation effects of large- and medium-sized cities to gradually expand the formation of regional economic belts and continue to improve the road network planning. Based on the economic development of the central and western regions, pay attention to the construction of high-speed railway infrastructure in the central and western regions. For busy areas, consider increasing the auxiliary railway network.

As shown in Figure 14, in order to better present the simulation results, the calculation example parameters are set on the basis of real data. Refer to the national highway odometer, 12306 official website of high-speed railway operating mileage, and China Southern Airlines flight mileage for the node distance under different transportation modes. Since the transit time of the node gradually increases according to the transportation process, the setting of the

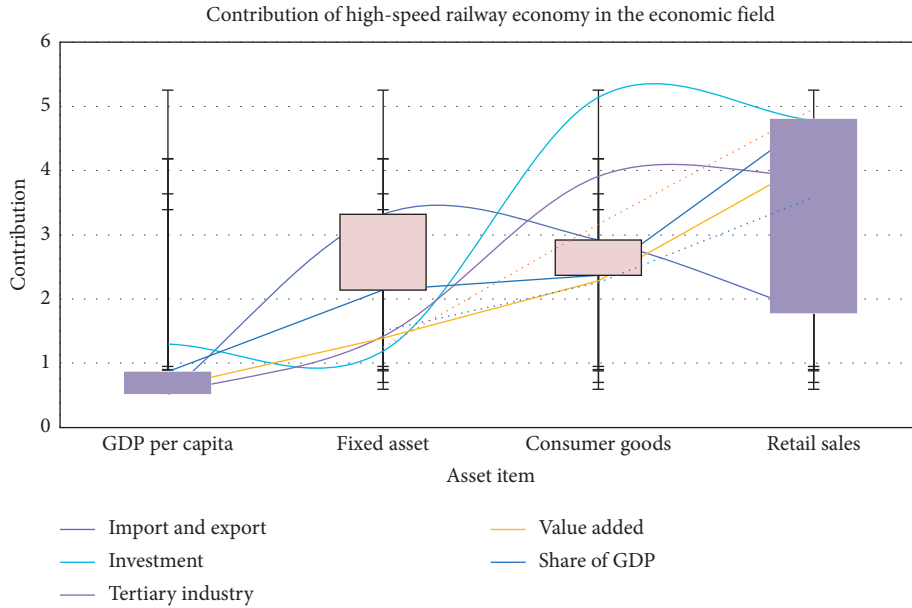


FIGURE 11: Contribution of high-speed railway economy in the economic field.

TABLE 13: High-speed railway economic data in the economic field.

| Item           | Import and export | Investment | Tertiary industry | Value added | Share of GDP |
|----------------|-------------------|------------|-------------------|-------------|--------------|
| GDP per capita | 0.53              | 1.3        | 0.52              | 0.63        | 0.87         |
| Fixed asset    | 3.32              | 1.19       | 1.42              | 1.39        | 2.14         |
| Consumer goods | 2.92              | 5.14       | 3.91              | 2.29        | 2.37         |
| Retail sales   | 1.78              | 4.78       | 3.92              | 4.16        | 4.8          |
| Investment     | 3.79              | 4.31       | 2.22              | 2.28        | 4.35         |

TABLE 14: IBk algorithm risk assessment system analysis table of results.

|   | Mean  | Standard deviation | Standard error of the mean |
|---|-------|--------------------|----------------------------|
| Economic growth   | 1.49  | 11.624             | 1.965                      |
| Financial situation   | 1.42  | 11.523             | 1.923                      |
| Personnel employment and personnel salary by industry by region | -1.33 | 9.643              | 1.275                      |

node service time window follows a certain order of increase. The transportation unit price, transit time, transit cost, and time window data are shown in Table 16. Among them, the high-speed railway and aviation time window widths at the nodes are set to 0.5 and 1 h, respectively.

As shown in Figure 15 and Table 17, most of the previous researches on multimodal transportation routes tend to be in the ideal state of no node time window. If the alternative transportation mode has no service time window constraint at the node, there is no need to wait for the goods in the

transit process, and the transit efficiency will be improved, a substantial increase. Under the condition of keeping the assumptions and other parameter settings unchanged, the path model without node time window is solved, and the optimal path is obtained as follows: AEHNO, the transportation mode combination is [3, 1, 1, 1], and the total transportation distance is 3654 km, the total path time is 28.29 h, and the total cost is 31057.38 yuan. Analyze and output the five path schemes with the smallest total cost, as shown in Table 5.

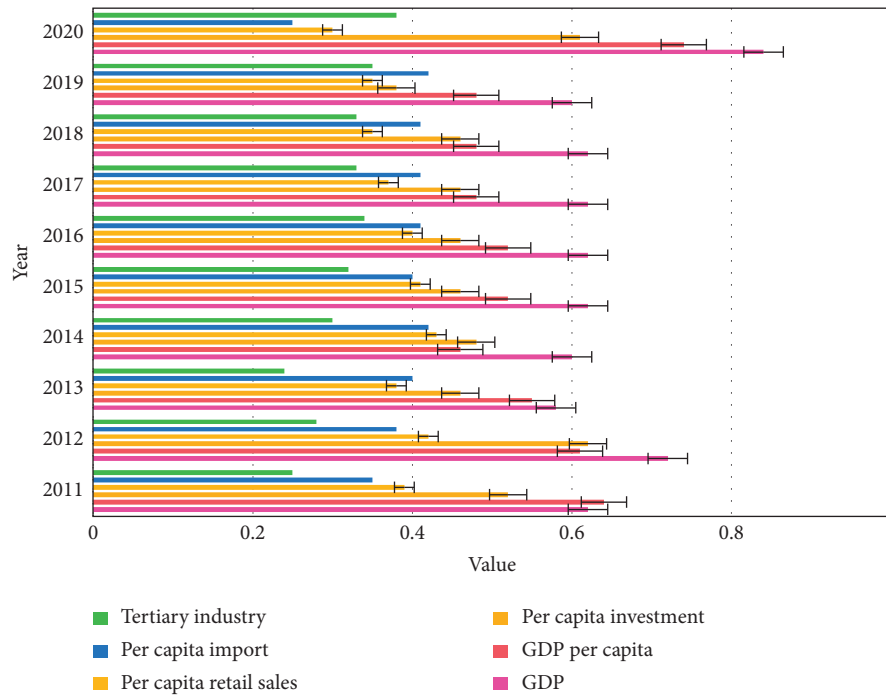


FIGURE 12: IBk algorithm-based risk assessment system to analyze the results.

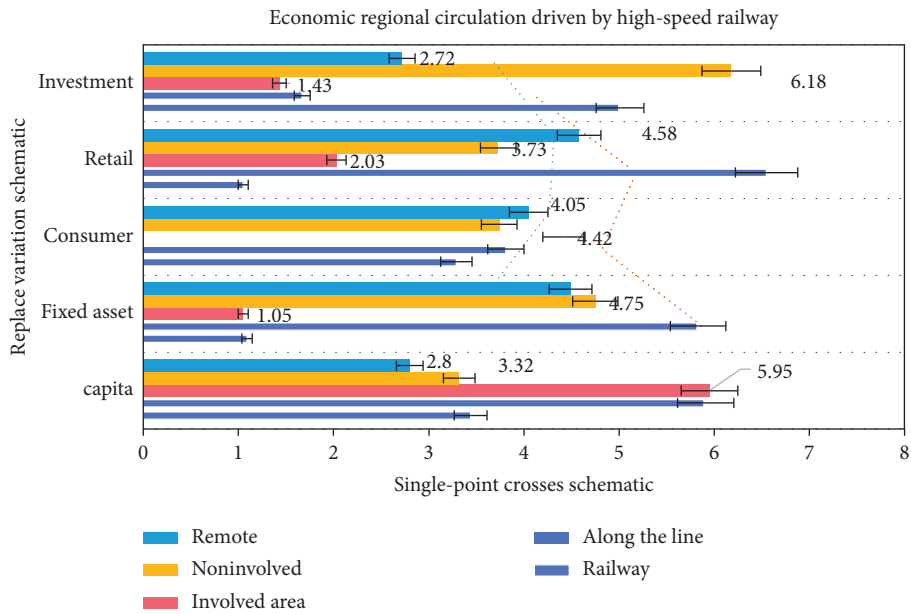


FIGURE 13: Economic regional circulation driven by high-speed railway.

TABLE 15: Economic area circulation data.

| Item        | Railway | Along the line | Involved area | Noninvolved | Remote |
|-------------|---------|----------------|---------------|-------------|--------|
| Capita      | 3.44    | 5.91           | 5.95          | 3.32        | 2.8    |
| Fixed asset | 1.09    | 5.83           | 1.05          | 4.75        | 4.49   |
| Consumer    | 3.29    | 3.81           | 4.42          | 3.74        | 4.05   |
| Retail      | 1.05    | 6.55           | 2.03          | 3.73        | 4.58   |
| Investment  | 5.01    | 1.67           | 1.43          | 6.18        | 2.72   |



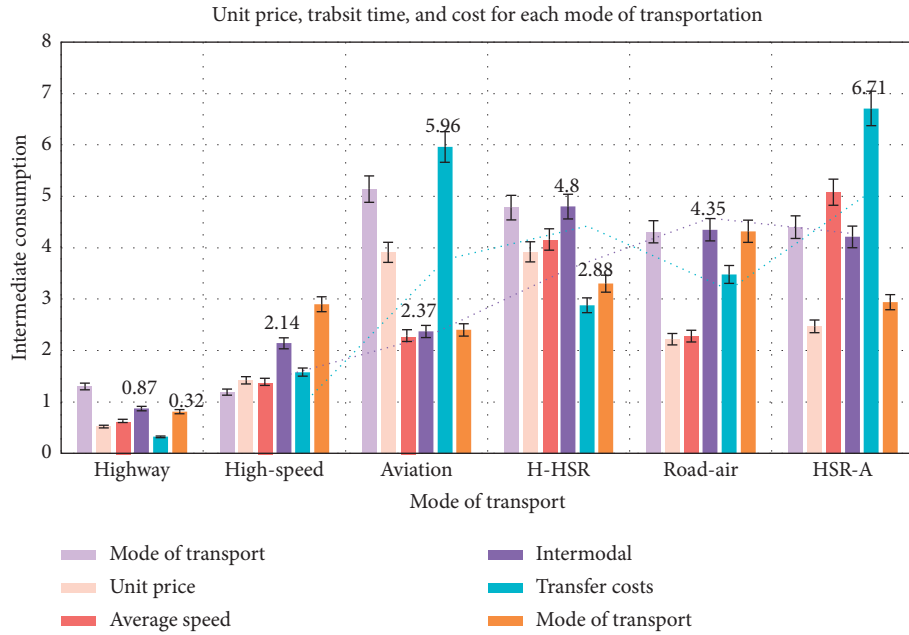


FIGURE 14: Unit price, transit time, and cost for each mode of transportation.

TABLE 16: Unit price, transit time, and cost for each mode of transportation.

| Mode of transport           | Mode of transport | Unit price | Average speed | Intermodal | Transfer costs | Mode of transport |
|-----------------------------|-------------------|------------|---------------|------------|----------------|-------------------|
| Highway                     | 1.3               | 0.52       | 0.63          | 0.87       | 0.32           | 0.81              |
| High-speed railway          | 1.19              | 1.42       | 1.39          | 2.14       | 1.58           | 2.9               |
| Aviation                    | 5.14              | 3.91       | 2.29          | 2.37       | 5.96           | 2.4               |
| Highway-high                | 4.78              | 3.92       | 4.16          | 4.8        | 2.88           | 3.3               |
| Road-air                    | 4.31              | 2.22       | 2.28          | 4.35       | 3.48           | 4.32              |
| High-speed railway-aviation | 4.4               | 2.47       | 5.08          | 4.21       | 6.71           | 2.94              |

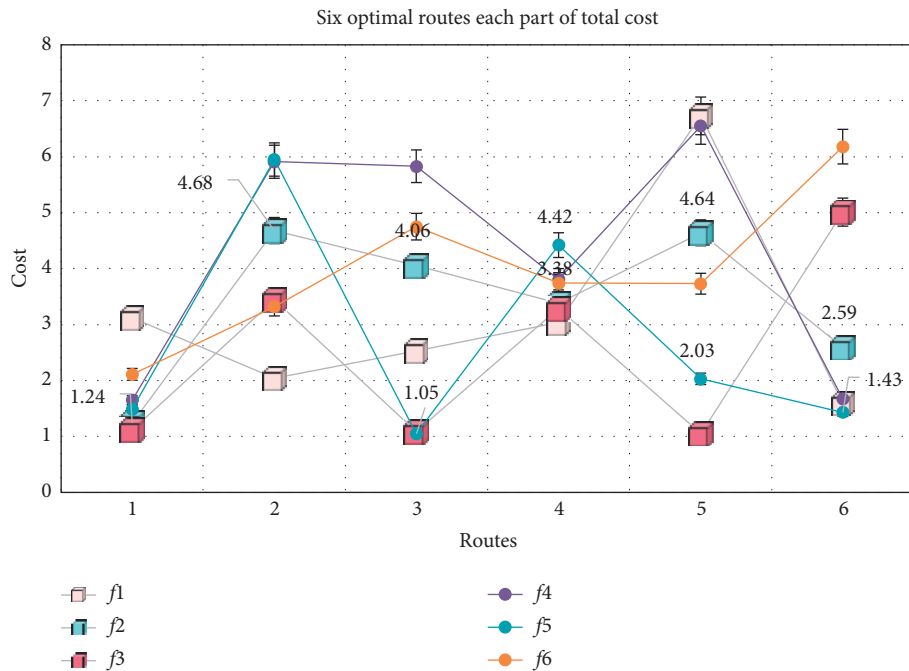


FIGURE 15: Six optimal routes of each part of total cost.

TABLE 17: 6 optimal routes of each part of total cost.

| Num. | $f_1$ | $f_2$ | $f_3$ | $F_4$ | $f_5$ | $f_6$ |
|------|-------|-------|-------|-------|-------|-------|
| 1    | 3.12  | 1.24  | 1.12  | 1.65  | 1.49  | 2.11  |
| 2    | 2.04  | 4.68  | 3.44  | 5.91  | 5.95  | 3.32  |
| 3    | 2.53  | 4.06  | 1.09  | 5.83  | 1.05  | 4.75  |
| 4    | 3.04  | 3.38  | 3.29  | 3.81  | 4.42  | 3.74  |
| 5    | 6.73  | 4.64  | 1.05  | 6.55  | 2.03  | 3.73  |
| 6    | 1.59  | 2.59  | 5.01  | 1.67  | 1.43  | 6.18  |

## 5. Conclusions

The data storage system ensures the timeliness and sharing of inventory data. In the past, data inventories used inventory personnel to access and manually import electronic devices. The existing system can meet the direct input of the electronic inventory unit, and the inventory staff can view and modify it in real time, thereby saving time; at the same time, the inventory data network can allow all levels from the top to the bottom to control and process the data to avoid a situation that the superior had to submit long data to subordinates to understand the data status and the system to perform data exchange. The intelligent data analysis system uses the current popular data mining technology. National financial inventory data meets a number of characteristics. Although it has some possible laws, there are inevitably many inherent problems, which are difficult to find. Generally speaking, big data has broad prospects for the development of information services in integrated transportation hubs, which need to be explored.

Because the traditional economy cannot conduct comprehensive research on various factors that affect the economy, this article analyzes data mining technology in a targeted manner to find a data mining algorithm suitable for regional economic analysis. Enhancing the accessibility of the railway transportation system will bring about huge economic benefits to the region, and improving the economic benefits will enable the country to increase the construction of railway transportation facilities to better serve economic development and improve the relationship between the two. Development is also an important way to promote the sustainable and healthy development of cities. Strengthening the importance of coordination between the two and studying the interaction between railway accessibility and regional economic development are important research topics for China's railway development.

There is a two-way feedback mechanism between the railway transportation system and the regional economic system. Railway transportation mainly affects the regional economy from two aspects: the overall economy and the spatial industrial structure. The investment in railway construction will affect the expansion of the regional structure and provide new impetus for economic development. On the contrary, the development of the regional economy will have a positive impact on railway transportation from three aspects: the supply of construction funds, the demand for railway transportation, and the supply of labor and technology. The survey results show that, in order to maintain the coordination between railway transportation and regional economic development, the state

must control the investment in railway construction to achieve a virtuous circle, narrow the gap between regions, and improve the level of coordination.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] L. Li, A. Tsunekawa, I. Maclachlan, G. Li, A. Koike, and Y. Guo, "Conservation payments, off-farm employment and household welfare for farmers participating in the "grain for green" program in China," *China Agricultural Economic Review*, vol. 12, no. 1, pp. 71–89, 2019.
- [2] B. Wang and B. P. Y. Loo, "Travel time use and its impact on high-speed-railway passengers' travel satisfaction in the e-society," *International Journal of Sustainable Transportation*, vol. 13, no. 5, pp. 197–209, 2019.
- [3] J.-P. Pascal and B. Soua, "Solving conformal contacts using multi-Hertzian techniques," *Vehicle System Dynamics*, vol. 54, no. 6, pp. 784–813, 2016.
- [4] K. Ikeda, K. Murota, T. Akamatsu et al., "Agglomeration patterns in a long narrow economy of a new economic geography model: analogy to a racetrack economy," *International Journal of Economic Theory*, vol. 13, no. 1, pp. 113–145, 2016.
- [5] Y. Han, Z. Geng, Y. Qu, and Q. Zhu, "Linear optimization fusion model based on fuzzy C-means: case study of energy efficiency evaluation in ethylene product plants," *Journal of Analytical and Applied Pyrolysis*, vol. 125, no. 5, pp. 347–355, 2017.
- [6] H.-Q. Li, X.-F. Liu, S.-J. Guo, and G.-P. Cai, "Deployment dynamics and control of large-scale flexible solar array system with deployable mast," *Advances in Space Research*, vol. 58, no. 7, pp. 1288–1302, 2016.
- [7] W. Xia and A. Zhang, "High-speed railway and air transport competition and cooperation: a vertical differentiation approach," *Transportation Research Part B: Methodological*, vol. 94, no. 11, pp. 456–481, 2016.
- [8] Y. Bi, C. Lin, H. Zhou, P. Yang, X. Shen, and H. Zhao, "Time-constrained big data transfer for SDN-enabled smart city," *IEEE Communications Magazine*, vol. 55, no. 12, pp. 44–50, 2017.
- [9] Y. Qin and L. Jia, "Train reliability and safety analysis," *Active Safety Methodologies of Rail Transportation*, vol. 10, no. 4, pp. 119–165, 2019.

- [10] K. Sato and Y. Chen, "Analysis of high-speed railway and airline transport cooperation in presence of non-purchase option," *Journal of Modern Transportation*, vol. 26, no. 4, pp. 5–28, 2018.
- [11] S. Lin, Q. Yu, Z. Wang et al., "A fault prediction method for catenary of high-speed rails based on meteorological conditions," *Journal of Modern Transportation*, vol. 27, no. 3, pp. 211–221, 2019.
- [12] T. Feng and S. Jungblut, "Foreign direct investments in the process of transition," *Economic Survey*, vol. 5, no. 2, p. 273, 2018.
- [13] S. Lan, C. Yang, and G. Q. Huang, "Data analysis for metropolitan economic and logistics development," *Advanced Engineering Informatics*, vol. 32, no. 4, pp. 66–76, 2017.
- [14] G. Qorraj, "Towards European union or regional economic area: western balkans at crossroads," *Naše Gospodarstvo/Our Economy*, vol. 64, no. 1, pp. 11–17, 2018.
- [15] M. Lili and J. Tianying, "68. Study on spatial-temporal differences in coupling coordination between urbanization, regional innovation and economic growth in China," *Boletín Técnico/Technical Bulletin*, vol. 55, no. 10, pp. 480–489, 2017.
- [16] Y. Chen, H.-L. Chang, and C.-W. Su, "Does real wage converge in China?" *Journal of Economic Interaction and Coordination*, vol. 11, no. 1, pp. 77–93, 2016.
- [17] L. Han, "Labor market changes during transformation and upgrading of China's three major special economic zones," *Asian Agricultural Research*, vol. 10, no. 10, pp. 5–10, 2018.
- [18] G. Sargent, K. R. Perez-Daniel, A. Stoian et al., "A scalable summary generation method based on cross-modal consensus clustering and OLAP cube modeling," *Multimedia Tools and Applications*, vol. 75, no. 15, pp. 9073–9094, 2016.
- [19] T. Yuan, W. Deng, J. Hu, Z. An, and Y. Tang, "Unsupervised adaptive hashing based on feature clustering," *Neurocomputing*, vol. 323, no. 5, pp. 373–382, 2019.
- [20] H. Shibuya and S. Maeda, "Anomaly detection method based on fast local subspace classifier," *Electronics and Communications in Japan*, vol. 99, no. 1, pp. 32–41, 2016.
- [21] H. T. Zheng, W. Zhe, W. Wei et al., "Learning-based topic detection using multiple features," *Concurrency and Computation: Practice and Experience*, vol. 30, no. 15, pp. 1–11, 2018.
- [22] F. Zhang, Y. Zhang, H. Yuan et al., "Analysis and research of government procurement audit in universities based on big data environment," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 21, no. 2, pp. 457–464, 2018.
- [23] S. Yu, "Research on the early warning of college graduates based on big data," *Revista de la Facultad de Ingeniería*, vol. 32, no. 12, pp. 121–128, 2017.
- [24] Y. Zhao, "Manufacturing personalization models based on industrial big data," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 21, no. 6, pp. 1287–1292, 2018.
- [25] H. Li, X. Zhou, J. Shen, and D. Luo, "Analysis on the issues in ISO 6892-1 and TENSTAND WP4 report based on the data of confirm tests by 21 laboratories," *Journal of Testing and Evaluation: A Multidisciplinary Forum for Applied Encees and Engineering*, vol. 45, no. 3, pp. 1105–1114, 2017.
- [26] A. V. Shcherbakov, V. P. Rubtsov, A. S. Kozhechenko, and M. V. Ivashchenko, "Development of standardized circuit solutions for control and data acquisition systems of a research complex on the basis of electron-beam technological installations," *Russian Electrical Engineering*, vol. 87, no. 6, pp. 340–346, 2016.
- [27] U. Nazarbek, U. Besterekov, S. Nazarbekova, O. Beisenbayev, and A. Bolysbek, "Obtaining complex mineral fertilizers on the basis of a technogenic wastes of phosphorus production," *Oriental Journal of Chemistry*, vol. 32, no. 4, pp. 2027–2033, 2016.

## Research Article

# Museum Display Showcase Furniture System Research Based on Internet of Things Technology in Intelligent Environment

Jiaojiao Hu,<sup>1</sup> Zhihui Wu ,<sup>1</sup> and Lei Jin<sup>2</sup>

<sup>1</sup>College of Furnishings and Industrial Design, Nanjing Forestry University, Nanjing 210037, China

<sup>2</sup>Zhejiang Art Museum, Hangzhou 310000, China

Correspondence should be addressed to Zhihui Wu; wzh550@njfu.edu.cn

Received 22 October 2020; Revised 26 December 2020; Accepted 4 January 2021; Published 29 January 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Jiaojiao Hu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The protection of cultural relics has always been an important issue in the field of museums and archaeology. With the development of Internet of Things technology, the security system of the museum is more intelligent and integrated. In order for the museum display system to keep up with the intelligent age, this article mainly studies the research and realization of the museum showcase system based on the Internet of Things technology in a smart environment. Before the start of the experiment, we developed the overall design of the system, including three functional modules: temperature module, illumination module, and monitoring module. The experiment is done mainly for system testing. The performance test of the sensor module needs to sample the temperature and humidity sensor to verify the accuracy of the temperature and humidity signal collected in the instrument circuit. The light information collection test uses the ADC sampling inside the CC2530 to obtain the required data and judges whether these temperature, humidity, and light intensity values exceed the preset values. The serial port needs to be initialized to carry out data communication and transmission normally. After the receiving end finishes receiving, the sending end will clear the data buffer to prepare for the next data transmission. The experimental data show that the error between the predicted value and measured value of the temperature system is about 3°C, which is within the allowable error range of the experiment. The results show that the system has perfect functions, is safe and reliable, meets the expected requirements, and has good practicability.

## 1. Introduction

The Internet of things is a new network communication technology in recent years. It is a new network based on the Internet, through various information sensing devices, using cloud computing and big data technology, to connect “human machine things” together by means of communication, so as to realize informatization and intelligence. With the development of social economy, people pay more and more attention to the construction of spiritual civilization. Museums play an important role in providing social education, recording historical changes, and displaying cultural details. It has become an indispensable part of the construction of human spiritual civilization. As an important part of the museum, the display cabinet has a close relationship with the exhibits, so the design of the

display cabinet has a strong influence on the overall exhibition effect.

Cultural relics are the product of history. In the course of five thousand years of development, the Chinese nation has created brilliant material and spiritual civilizations for us but also left us countless treasures of cultural relics. Only by better preserving these cultural heritages in the form of material and promoting them to better serve the development of mankind can it promote the better development of mankind. As the highly integrated and comprehensive application of a new generation of information technology, the rapid development of the Internet of Things plays an important role and significance in promoting the design of museum showcases and the green, intelligent, and sustainable development of the

society. According to the specific environment, the corresponding sensor is selected to obtain the parameters, which improves the accuracy of obtaining the environmental parameters and eliminates the drawbacks of manual detection.

The Internet of Things technology has brought tremendous changes to human life. Wu believes that, in recent years, museums have invested a lot of manpower and material resources to promote the construction of digital museums. However, in the construction of digital museums, it is very difficult to manage so many collections and find a collection in them. Therefore, he used information technology to identify collections. He described the digital construction of the museum and proposed the idea of using the Internet of Things technology in the construction of the digital museum. He proposed a system design for museum collection management, which provides a new way for the construction of smart museums. Although his research is theoretically correct, it lacks specific experimental content [1]. Palattella believes that, in recent years, a variety of communication technologies have gradually emerged, reflecting the diversity of application fields and communication requirements. He analyzed in detail the potential of 5G technology in the Internet of Things by considering technology and standardization. He reviewed the current IoT connectivity landscape and the main 5G support factors for IoT. He explained that the close connection between the Internet of Things and 5G may cause huge business changes in the operator and supplier ecosystem. His research is not comprehensive enough [2]. Bello believes that similar to the way humans use the Internet, devices will become the main users of the Internet of Things (IoT) ecosystem. Therefore, device-to-device (D2D) communication is expected to become an inherent part of the Internet of Things. Devices will automatically communicate with each other without any centralized control and cooperate in a multihop manner to collect, share, and forward information. He believes that the ability to collect relevant information in real time is the key to harnessing the value of the Internet of Things because such information will be transformed into intelligence, which will help create a smart environment. Ultimately, the quality of the information collected depends on the intelligence of the device. He outlined how to implement smart D2D communication in the IoT ecosystem. His research is not accurate enough [3]. Akpakwu believes that the Internet of Things (IOT) is a promising technology, which tends to change and connect the global world through seamless connection and heterogeneous intelligent devices. The current demand for machine-type communication (MTC) has led to a variety of communication technologies and services to achieve the vision of modern IOT. He surveyed the latest Internet of things application requirements and related communication technologies. In addition, he discussed in detail the 3G partner program's cellular-based low-power wide area solution to support and enable new service requirements for "large to critical IOT" use cases. He presented a comprehensive overview of emerging technologies and enabling technologies, with a focus on 5G mobile networks, to support the exponential traffic growth of IOT. Although his research is comprehensive, it lacks necessary data [4].

The innovation of this article is to connect this emerging wireless communication technology with the current cultural relics protection hotspot and design a museum showcase system based on the Internet of Things technology. The system can monitor environmental information parameters in real time and realize the monitoring of multiple environmental parameters in the same network. It can not only monitor the environmental parameters that affect the preservation of cultural relics in the showcase in real time but also adjust the temperature and humidity in the showcase in real time. Compared with the traditional environmental monitoring system, this scheme has good scalability, convenient use, and strong portability. It can be used for environmental monitoring in other fields through simple parameter modification. At the same time, this paper compares and analyzes the key technologies in the system and focuses on the detailed introduction and analysis of single-chip technology, WiFi technology, cloud computing mode, and other technologies. Finally, the overall architecture of the system is built, the design plan is determined, and the appropriate chip is selected for design.

## 2. Internet of Things and Showcase System

*2.1. Internet of Things Technology.* Information and communication technology currently has the ability to transmit any place and any information to designated people or objects in a short time, and this information line is connected into an intricate network, which is the Internet of Things. The main function of the Internet of Things technology is to finally realize a technology of information transmission between objects through the perception and transmission of information. The roles and functions of the various structures of the Internet of Things technology are also different. The perception layer is mainly used for the perception of information, such as information collection and recognition; the network layer is mainly used for the transmission of various information, and the application layer is mainly presents the result of the final identification information to humans. The result of the final identification information is presented in front of human beings, which further promotes human's intelligent management of something [5]. Sensor nodes generally belong to embedded systems (micro) in the aforementioned Internet of Things, and their basic functions are mainly manifested in routing, wireless communication, and collecting, storing, and processing data. In addition to collecting and processing local information, it is also necessary to forward and save data transmitted from other nodes. The electrical energy of the sensor node mainly comes from the battery, so its energy consumption is positively related to its function. The more nodes in the Internet of Things, the more data will need to be forwarded by sensor nodes, which will increase its energy consumption. The monitoring area of the heterogeneous Internet of Things is divided into multiple clusters. Each cluster is usually composed of several nodes (sensor nodes) in the cluster and one cluster head node (sink node). All sensor nodes in the cluster need to transmit data to the cluster aggregation node, and then, each cluster aggregation

node in the network will send the collected data to the task management node through satellite channels, the Internet, and other channels [6, 7].

The network layer is an indispensable part of the ZigBee protocol stack. It is connected with the media access control layer and the user layer confidentiality. The joining and leaving of nodes in the network, routing lookup, and other functions must be implemented through this layer. The network layer reference model is shown in Figure 1. The network layer provides a function interface to the media access control layer to ensure the normal operation of this layer.

The source node forwards the data packet to its corresponding neighbor node, and then, it uses this routing algorithm to check that the destination node is its parent node. The parent address of the neighbor node is given as follows:

$$P_{LA} = L_A + \left[ \frac{D_A - (L_A + 1)}{C_{\text{skip}}(d)} \right] \times C_{\text{skip}}(d) + 1. \quad (1)$$

The Euclidean distance between two points is given as follows:

$$d(x_i, x_j) = \sqrt{\left( |x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \cdots + |x_{in} - x_{jn}|^2 \right)}. \quad (2)$$

Among them,  $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$  and  $x_j = (x_{j1}, x_{j2}, \dots, x_{jn})$  are the data objects of the monitoring points.

The formula for the average distance between monitoring data points is given as follows:

$$\text{avgDist} = \frac{1}{c_N^2} \times \sum d(i, j). \quad (3)$$

In the formula,  $c_N^2$  is the number of combinations of two points taken from  $n$  points.

The square error criterion function is as follows:

$$\text{SSE} = \sum_{i=1}^k \sum_{m \in c_i} |m - m_i|^2. \quad (4)$$

The difference within the cluster measures the compactness within the cluster, that is, the density distribution of the data points in the museum. It can be judged here that the sum of the squares of the distance from each data point in the cluster to the cluster center is calculated. The formula is as follows:

$$w(c) = \sum_{i=1}^k w(C_i) = \sum_{i=1}^k \sum_{x \in c_i} d(x, C_i)^2. \quad (5)$$

The balanced environmental evaluation function is as follows:

$$W(c, k) = \sqrt{w^2(c) + b^2(c)}. \quad (6)$$

From  $t=1$  to  $t=T$ , in the 16 channels of the ZigBee wireless sensor network, the probability of the hidden state observation sequence is given as follows:

$$\Pr(x^T) = \prod_{t=1}^T a_{x_{t-1}x_t}, \quad (7)$$

where  $x_t$  is the hidden channel state of the channel at time  $t$ . Under the given hidden state sequence  $x^T$ , from  $t=1$  to  $t=T$ , the conditional probability of the observation state sequence  $y^T$  is given as follows:

$$\Pr(y^T | x^T) = \prod_{t=1}^T b_{x_t}(y_t). \quad (8)$$

In the KSS-HMM model, the historical hidden state and detected channel state are determined. Therefore, if the ZigBee network device performs channel switching at time  $t_0$ , the probability of the observation sequence of the channel time slot time  $T_1$  is given as follows:

$$\Pr(y^{T_1}) = \prod_{t=t_0-T_1+1}^{t_0} a_{x_{t-1}x_t} b_{x_t}(y_t), \quad (9)$$

where  $y_1^{T_1}$  is the observation state sequence of the time slot time  $T_1$ .

Based on the known observation sequence  $y_1^{T_1}$ , the expected channel idle time in the future is given as follows:

$$E(T) = \sum_{T_2=1}^{\infty} T_2 \cdot \Pr(y_1^{T_1}, y_2^{T_2}). \quad (10)$$

**2.2. Museum Showcase.** Museum display cabinet refers to the cabinet specially designed and manufactured by the museum for the storage and display of collections to meet the needs of collection display and realize the purpose of research, education, and appreciation. From the definition of museum display cabinet, a museum display cabinet must serve as the display of the museum and realize the purpose of museum design and production. The lighting mode of the independent cabinet generally adopts top-line fluorescent lamp lighting or spotlight lighting. In order to prevent the damage of lighting lamps to cultural relics, the lighting part and display part are generally separated by aluminum grille and glass [8]. The function of the aluminum grille is to provide uniform illuminance so that visitors cannot see the structure of the lamp. The upper part of the grille is generally made of glass with an antiultraviolet film to effectively isolate harmful light and heat. The display cabinets of precious cultural relics generally adopt optical fiber lighting. This lighting method generates less heat and can protect cultural relics to the utmost extent. The full-body wall cabinet is generally placed against the wall, usually used to display large or grouped cultural relics, and the display part usually has side panels or back panels, which can only be viewed from the front or from both sides. The display cabinets of precious cultural relics generally adopt optical fiber lighting. This lighting method generates less heat and can protect cultural relics to the utmost extent. The lighting of table-style showcases generally uses external lighting, and some precious paper or organic cultural relics use optical fiber

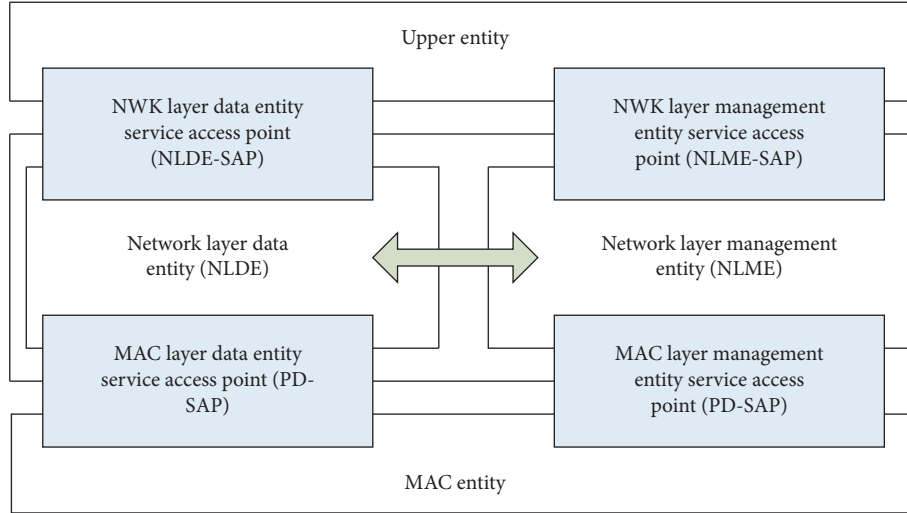


FIGURE 1: Network layer reference model.

lighting in the cabinet. This lighting method generates less heat and can protect cultural relics to the utmost extent [9, 10].

PID controller can be expressed as follows:

$$u(t) = k_p \left( \text{err}(t) + \frac{1}{T_i} \cdot \int \text{err}(t) dt + \frac{T_D d_{\text{err}(t)}}{dt} \right). \quad (11)$$

Among them,  $k_p \text{err}(t)$  is a proportional control item.

The PI mathematical model is given as follows:

$$\mu(t) = K_p \left[ e(t) + \frac{1}{T_I} \int_0^t e(t) dt \right] + \mu_0. \quad (12)$$

The transfer function form is given as follows:

$$G(s) = \frac{U(s)}{E(s)} = K_p \left( 1 + \frac{1}{T_I s} + T_D s \right). \quad (13)$$

When increasing the gain, the system will become more sensitive. But when the gain is too large, the number of system vibrations will increase and the adjustment time will

be prolonged. Therefore, it is very important to determine the appropriate gain value in the control process.

**2.3. System Design.** To realize the monitoring of showcase parameters, corresponding sensors are needed. In addition to sensors, instruments are also needed to collect and process sensor data and can control humidity in a certain way and send the data to the host computer after networking [11].

**2.3.1. Temperature Module.** The momentum conservation equation follows Newton's second law of motion, and the differential form of conservation is expressed as follows:

$$\frac{\partial}{\partial t} (\rho u_i) + \frac{\partial}{\partial x_j} (\rho u_i u_j) = -\frac{\partial p}{\partial x_i} + \frac{\partial \tau_{ij}}{\partial x_j} + \rho g_i + F_i. \quad (14)$$

In the formula,  $p$  is static pressure, and  $F_i$  is unit mass force in different directions.

The energy conservation equation follows the first law of thermodynamics, and the differential form of conservation is expressed as follows:

$$\frac{\partial}{\partial t} (\rho E) + \frac{\partial}{\partial x_i} (u_i (\rho E + p)) = \frac{\partial}{\partial x_i} \left( k_{\text{eff}} \frac{\partial T}{\partial x_i} - \sum_j h_j J_j + u_j (\tau_{ij})_{\text{eff}} \right) + S_h. \quad (15)$$

In the formula,  $E$  is the total energy of fluid clusters, and  $k_{\text{eff}}$  is the effective thermal conductivity.

The system is mainly composed of the wireless sensor network and monitor platform. The zig node is composed of the zigs node and bee node. The ZigBee terminal node is connected with various sensors, which are used to collect environmental data in the museum. The TCP socket receiver program on the server side writes the data from GPRS DTU to the database on the server side and then uses the php script program based on Web technology to display the data to the

user. The user can analyze and process the data through the PC client [12]. The intelligent greenhouse communication mode needs to collect as much data as possible and also uses video surveillance. Therefore, the data rate of network transmission is relatively high to accommodate video communication. Transmission of information should be secure, and information security has been paid more and more attention now, and there will be no application market if there is no security application; since the equipment is the uninterrupted automatic work, the communication stability needs to be guaranteed [13, 14].

In order to measure the temperature and humidity value, a sensor is needed to measure it. This design uses the integrated temperature and humidity sensor DHT11. DHT11 has undergone precise humidity adjustment and stores the school team coefficients in the memory and calls these calibration coefficients during detection signal processing. Generally, a 4-pin package is used, in which the 3 pins are floating, and the second pin is transmitted by a single bus. The photoresistor module is generally used to detect the brightness of the light in the surrounding environment, trigger the single-chip microcomputer or relay module, etc [15].

The methods and steps to determine the humidity index of cultural relics preservation are summarized as follows: when there is no moisture absorption model corresponding to various materials, the moisture regain of materials under different humidities is measured by the experiment, and then, the moisture absorption isotherm is drawn according to the data obtained; the parameter values of the moisture absorption model are obtained by nonlinear regression of several existing moisture absorption models with 1<sup>st</sup>Opt statistical software. Analyze the fitting procedures of each moisture absorption model and material moisture absorption isotherm, and find the model with the highest fitting accuracy; the model with the highest fitting accuracy is the best moisture absorption model of the material. Through this model, the relative humidity index of the material can be deduced by giving the moisture regain value under the condition of known temperature [16].

**2.3.2. Illumination Module.** In terms of hardware, the CC2530 chip is selected to deploy the wireless sensor network, and the PWM module is used as an LED-dimming module; in terms of software, an IAR-integrated development environment is used as the development tool, MySQL is used as the database software, and easyUI framework is used to realize the function of data viewing on Web. The illuminance sensor adopts nhzd10au series. The special-purpose optical chip has the advantages of low illumination and strong illumination. Sensor measurement signal transmission distance is long, and signal stability and transmission anti-interference ability are strong [17]. The system consists of server, gateway, and intelligent lighting terminal. The server serves as a node for collecting and storing terminal information and provides users with an interface for registering, inquiring, and controlling terminal lighting equipment. WEB/APP server, database server, and terminal access server are set up on the server. The system consists of server, system gateway, and intelligent lighting terminal. The server is the node for collecting and storing terminal information and provides users with an interface for registering, querying, and controlling terminal lighting equipment. The server is equipped with Web/app server, database server, and terminal access server. The system gateway is responsible for maintaining and coordinating networking and exchanging communication between many terminals and servers, and the gateway is connected up to the

terminal access server. The smart lighting terminal is used to collect the status information of LED lights and control the LED status, which is the final execution system of user actions [18, 19].

**2.3.3. Monitoring Module.** The hardware block diagram of the security instrument in the museum's smart showcase security system is shown in Figure 2. The security instrument uses the ARM CORTEX-M3-based STM32F103ZET6 microcontroller as the core controller to design the hardware circuit, which mainly includes a power-supply module, a real-time clock module, four groups of sensor monitoring modules, a WIFI module, a GSM communication module, and a booth lifting device. The display cabinet security instrument has less need for buttons, so it adopts the independent button design. The hardware circuit of the independent button is simple and easy to operate. It can be directly connected to the IO port of the single-chip microcomputer to obtain button information, which is widely used in the industrial embedded field [20]. Aiming at the cultural relics on display in the independent cabinet, this paper designs a booth lifting device. When the security system is triggered, if the booth is at the upper limit of the stroke, it will automatically descend to the dark warehouse of the showcase, reducing the risk of cultural relic damage. At the same time, the administrator can also control the lifting of the showcase booth through the remote control. The booth lifting device is composed of four DC motors and auxiliary mechanical devices to further improve work efficiency and ensure the safety of cultural relics. The security instrument controls the forward rotation, reverse rotation, and stop rotation of the electric push rod through the corresponding circuit to realize the function of lifting the booth. The remote control sends up, down, and stop commands to the security instrument through the wireless sending and receiving module PT2262/2272 [21].

### 3. Function Test of Museum Showcase System

**3.1. Experimental Environment.** Before testing the system, a test outline should be made in accordance with the system requirement documents, and a feasibility analysis of the test outline should be performed, to design reasonable test cases and to perform the test. The experimental equipment are as follows: a computer, a network coordinator (FFD) and a terminal node (RFD), a temperature and humidity sensor, ground temperature and humidity sensor, simulator, GPRS DTU, error detection wizard, and an integrated IAR development environment. Both test units are equipped with two independent data transmission and network control buttons and are connected to an LCD screen to observe the experimental results. Among these, the termination node is also used as a control center for the smart home system, which is connected to the higher computer via the RS232 channel, and the ZigBee network and operation can be further observed via the above computer [22, 23].



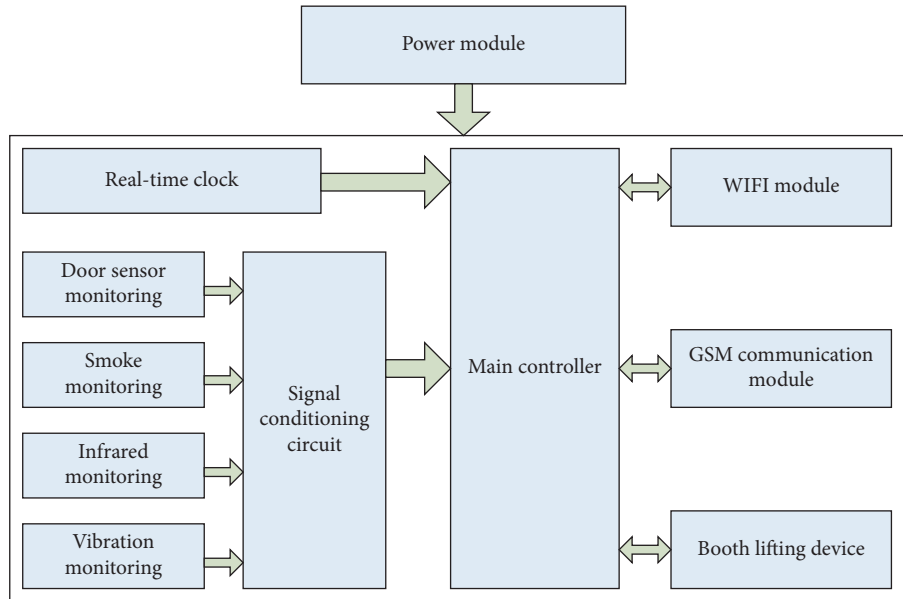


FIGURE 2: Hardware block diagram.

**3.2. System Function Test.** Functional testing is a verification test for each function of the system. According to the analysis of functional requirements and design requirements of the system, functional test cases are designed and tested item by item to check whether the prototype system has reached the functional goal of the design. Before performing functional testing, we must first build a testing environment. The steps to set up a test environment are as follows.

- (1) Open Eclipse, debug, and run the program, and start the system server.
- (2) Connect the Ethernet interface of the coordinator node to the router and the miniUSB interface to the 5 V power interface. The coordinator node starts to create the ZigBee network and waits for other nodes to join.
- (3) Deploy sensor nodes. After testing, the communication range of the node is 5–10 meters. Since the sensor is networked with a star structure and the coordinator node is placed in the museum, the nodes deployed in the museum can guarantee communication [24].

**3.3. Sensor Module Performance Test.** When collecting data on greenhouse environmental factors, in order to ensure the accuracy of its numerical reference, it is necessary to accurately test each sensor module. Various types of sensors are installed in different corners of the laboratory to ensure the integrity of data collection at any location in the laboratory. When detecting the received data, in order to be able to observe the change trend of the data more intuitively, select one of the sensors to collect data interface, compare the detection value obtained by the interface with the data collected by the actual engineering instrument, and verify the effectiveness of data collection [25]. By running the IAR

software tool, to various sensor modules of different types, the control instructions that conform to their own data acquisition process are programmed, which are the same principles as the operating instructions programmed in the coordinator module. When the coordinator module is called for communication, the LED1 diode can always keep on after each node applies to join the wireless network communication. A portable heating and humidifying device is used to gradually increase the temperature and humidity of the air around the room [26].

First, sample the temperature and humidity sensor to verify the accuracy of the temperature and humidity signal collected in the instrument circuit. Place the sensor and instrument in a built-up constant temperature and humidity room, set the temperature at 25°C and 40% RH, and use the serial debugging assistant to observe the collected data values. Then, the sensor whose signal type is analog signal is collected to verify that the Butterworth filter in the circuit will not make the data deviate, while ensuring the stability of the data. The debugging process is as follows: use a voltmeter to measure the voltage value connected to the AD pin of the microcontroller, use the serial debugging assistant to view the converted digital quantity, and judge the corresponding relationship between the two [27].

**3.4. Light Information Collection Test.** The collection of light intensity information is obtained by sampling the ADC inside the CC2530. It collects the amplified voltage signal of the BPW34S sensor and connects it to the port of the CC2530 chip. Finally, the sampled data is packaged and sent to the network coordinator in a wireless manner. The coordinator processes these data and judges whether these temperature, humidity, and light intensity values exceed the preset value. If it is greater than the previously set value, the

buzzer will send out an alarm signal so that the manager can turn on the corresponding button. The company's control equipment can adjust temperature, humidity, and light intensity [28].

*3.5. Serial Communication Function Test.* The serial port must be initialized before data communication, and transmission can be carried out normally. This is a prerequisite to ensure the smooth operation of the serial port. If the serial port wants to send data, first, save the data in the buffer of the sender and then start writing data through the serial port, and the data starts to transmit. After the receiver finishes receiving, the sender will clear the data buffer to prepare for the next data transmission [29].

## 4. System Test Results

*4.1. Temperature Module Test Results.* The exhibits in the museum must use a certain form of the museum showcase to achieve the purpose of the external display. At the same time, museum showcases must be based on exhibits and designed and manufactured in accordance with the different characteristics of different exhibits. Therefore, when discussing the design and selection of museum showcases, we must fully understand the relationship between exhibits and showcases, and on this premise, fully consider and study the impact of exhibits on showcases. The relationship between museum showcases and exhibits can be said to be interdependent and indispensable. On the one hand, in order to realize its display, education, and research functions, the museum needs to concentrate a large number of collections for the external display. In this way, it provides a broad stage and good prospect for the survival and development of the showcase. On the other hand, the showcase creates excellent conditions for the display of exhibits. It not only creates an extremely comfortable storage space for the exhibits inside but also effectively prevents external threats from harming the exhibits. In the museum's security system, it is the various security technologies and security equipment that ensure its survival and long-term maintenance. Various security technologies and security equipment play a huge role in the museum's daily security work.

The temperature test results are shown in Table 1 and Figure 3. In the control command test process, no matter it is single lamp control, area control, or global control, precise control is achieved. Whether it is time control, illuminance control, or instant control, real-time response is achieved, and there is no packet loss during communication. It is phenomenal that the control success rate is as high as 100%. From a macro point of view, the temperature and humidity environment in the cabinet will be affected by the changes in the environment outside the cabinet and show a relative change in time lag. The changes in the indoor environment outside the cabinet (exhibition room) will also be affected by the outdoor environment. The change of climate presents the effect of delay, and its delayed time-lag characteristics are generally affected by the construction

materials, airtight characteristics of the building, and display cabinet envelope structure, which presents different delay effects, and is vulnerable to the external environment. Under the influence of the external environment, the temperature change of the cabinet will show obvious fluctuations. The temperature change in the atmospheric environment has the regular characteristics of annual and daily changes. Similarly, the temperature environment in the exhibition room environment is not adjusted by the environmental adjustment equipment during the closing process. It will also show corresponding fluctuations, only because the time of the impact is lagging and the degree is weakened.

Simulation is carried out by NS2 software. In the simulation environment, the network range is  $200\text{ m} \times 200\text{ m}$ , and the number of network nodes is up to 220 and randomly distributed. The neighbor table of each node allows storage of 9 neighbor node information, and the network coordinator is located in the center of the network area. Considering the increasing number of nodes in the network, the end-to-end average delay is shown in Figure 4. The packet transmission rate is shown in Figure 5. The routing efficiency is shown in Figure 6. It can be seen from the simulation results that the improved ImcTR algorithm reduces the node and average energy consumption, while at the same time it also reduces the number of hops from the source node to the destination node and improves the network performance.

*4.2. Function Analysis of Showcase System.* The fuzzy control method requires that the equipment must be continuously adjustable or it can be divided into multiple levels with different control capabilities in order to design fuzzy control rules. In order to make fuzzy control have a certain adaptive ability, usually use the neural network or genetic algorithm for optimization. The regeneration of the dehumidifier mainly uses heating to evaporate the water in the solution. The temperature and humidity monitoring experiment data are shown in Table 2. The experimental data shows that the system test data is basically the same as the instrument test data, the temperature measurement error is  $\pm 0.3\%$ , and the humidity measurement error is  $\pm 3\%$ . During the experiment, the temperature of the simulated exhibition hall (constant temperature and humidity room) fluctuates around  $20^\circ\text{C}$ , and the humidity fluctuates around 45%. The initial temperature in the simulated showcase (standard test box) is  $20.0^\circ\text{C}$ , and the initial humidity is 40.6% RH. The initial temperature and humidity of the test box is different from that in the constant temperature and humidity room because of the tightness of the test box. Different positions, forms, specifications, and outlet wind speeds of air supply and exhaust ports will cause different indoor air flow conditions, and different air flow conditions will achieve different air conditioning effects. Reasonable airflow organization can make the personnel working area and specific area reach the required four degrees, namely, temperature, humidity, speed, and cleanliness, and its distribution is uniform and stable so that it can meet the requirements of human comfort and special production technology.

TABLE 1: Temperature test results.

|                      |        |        |        |        |        |        |        |        |        |        |        |        |
|----------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Single light control | 54.2 % | 51.6 % | 70.4 % | 68.5 % | 50.4 % | 63.1 % | 61.5 % | 67.4 % | 76.3 % | 45.3 % | 66.3 % | 41.0 % |
| Zone control         | 88.8 % | 63.3 % | 86.8 % | 76.9 % | 90.2 % | 57.7 % | 79.9 % | 92.1 % | 81.3 % | 90.5 % | 84.1 % | 74.8 % |
| Global control       | 98.2 % | 67.3 % | 87.4 % | 78.7 % | 72.3 % | 86.6 % | 70.4 % | 96.2 % | 67.7 % | 81.4 % | 74.2 % | 83.7 % |

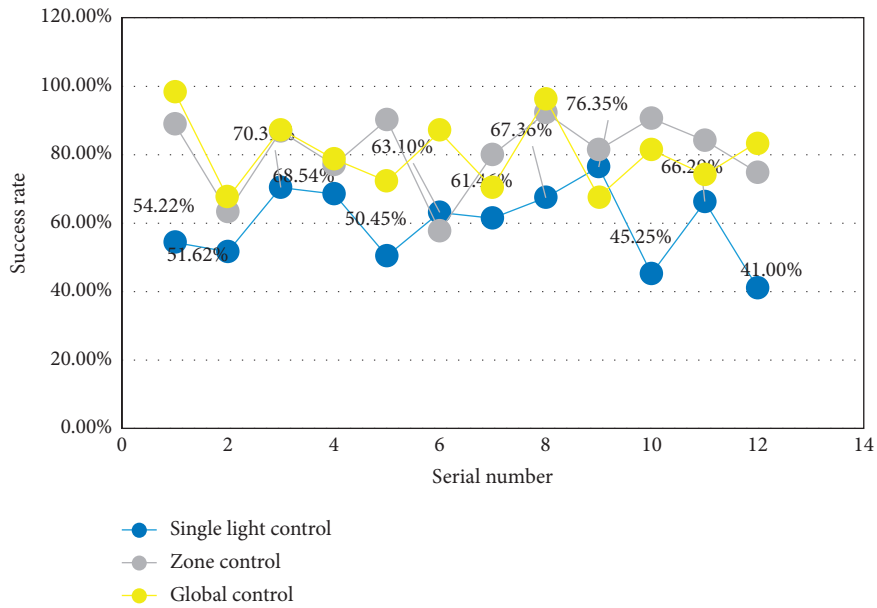


FIGURE 3: Temperature test results.

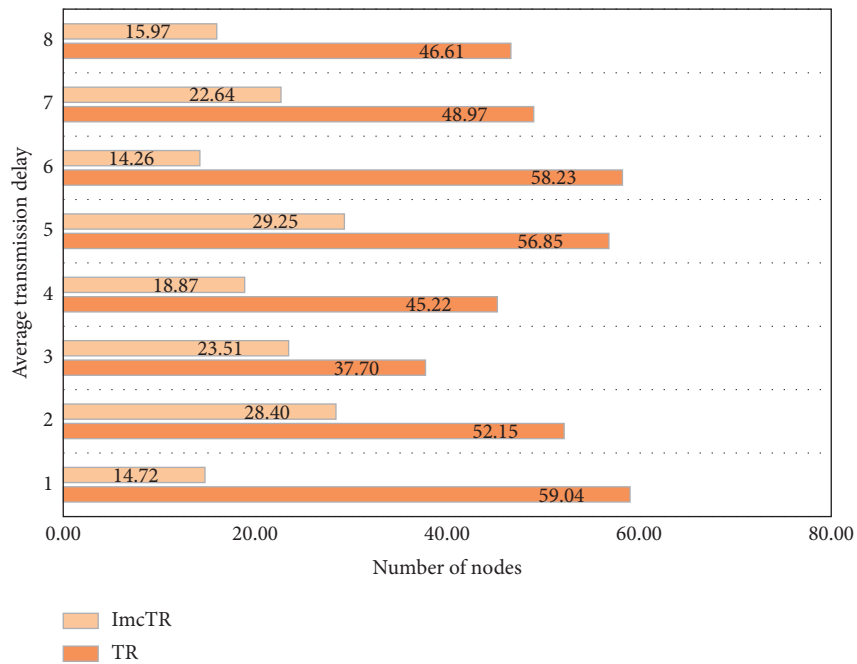


FIGURE 4: End-to-end average delay.

4.3. *Sensor Error Analysis.* System data collection: its collection method is the fusion of cluster collection and distributed collection. Distributed collection is to use the connection points of multiple sensing devices to lay out the

agent items of distributed collection in it to achieve the goal of data collection. Among them, the sensor device connection point transmits the collected information to the information-processing module through the wireless

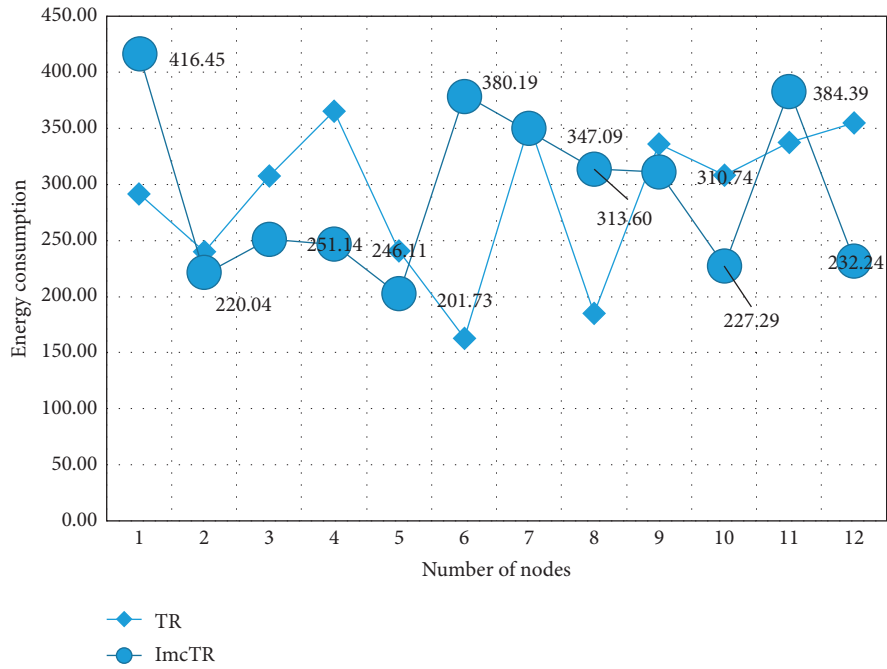


FIGURE 5: Packet transfer rate.

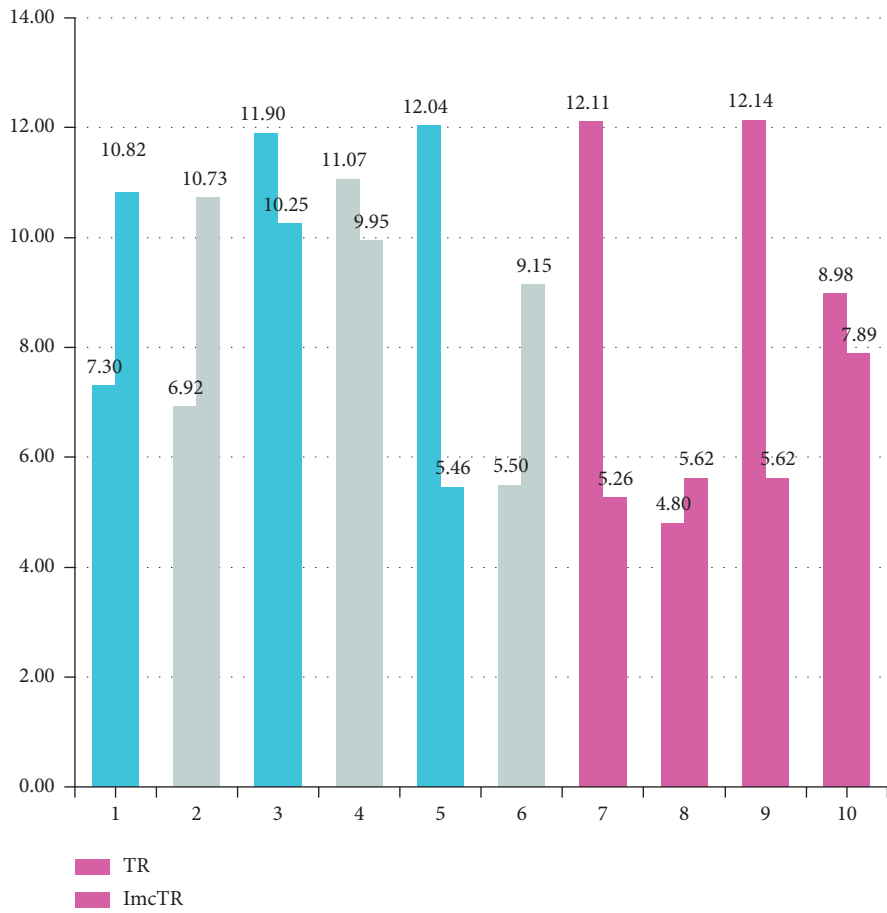


FIGURE 6: Routing efficiency.

TABLE 2: Comparison of temperature and humidity monitoring experimental data.

| Serial number | Standard temperature | Experimental temperature | Standard humidity | Experimental humidity |
|---------------|----------------------|--------------------------|-------------------|-----------------------|
| 1             | 16.7                 | 16.5                     | 45.2              | 45.1                  |
| 2             | 16.8                 | 16.6                     | 44.1              | 44.2                  |
| 3             | 17.3                 | 17.2                     | 42.8              | 43.0                  |
| 4             | 17.6                 | 17.6                     | 42.0              | 42.3                  |
| 5             | 18.0                 | 18.0                     | 41.2              | 41.3                  |
| 6             | 17.9                 | 17.8                     | 40.3              | 40.5                  |
| 7             | 18.3                 | 18.1                     | 40.0              | 40.1                  |
| 8             | 18.9                 | 19.0                     | 39.7              | 39.7                  |
| 9             | 19.5                 | 19.4                     | 39.6              | 39.7                  |
| 10            | 19.5                 | 19.5                     | 39.6              | 39.6                  |

network, which processes the data and at the same time feeds the processing results back to the upper computer client and loads the records in the database. Convergent collection is to realize the networking of multiple sensor equipment connection points, set up observation ports on the upper computer terminal, realize centralized observation of multiple sensor equipment connection points, and collect data information corresponding to the observation items. Data transmission is a process of transferring data from the information source to the database through one or several data links based on certain regulations. Its function is to exchange information among multiple nodes, and the information transmission methods are usually wired and wireless. Comparing the two, the wireless method has excellent adaptability, and the regional dimension has little influence. Before information enters the data fusion process, the step of data cleaning is essential. It is to delete invalid information and error information and duplicate information collected by the sensor to ensure the authenticity and validity of the data in the later fusion. Supplemented with the adaptive weighting algorithm, it can solve the time-varying and limited nature of independent sensing devices. The temperature data comparison is shown in Figure 7. It can be seen that there is a time interval between the measured values, indicating that the trend of the predicted value and measured value are close. The error between the predicted value and measured value is about 3°C, which is within the allowable error range of the experiment.

*4.4. System Accuracy Analysis.* In the design of relational database, the most important part of the work is to allocate metadata to various relational data tables and complete the classification of these metadata, and the operation of data will depend on the relationship between these data tables; through the relationship between these data tables, these data can be linked together in a meaningful way. In a relational database, creating a relationship between tables means that columns in one table are linked to columns in another table. This relationship can prevent redundant data and ensure that the information in one table matches that in the other. Relationships are generated by matching data in key columns, which are columns with the same name in two

tables. Usually, primary keys in one table are matched with external keys in another table. There are many kinds of harmful gases in the microenvironment of museum's cultural relics. When the cultural relics are exposed to a certain concentration of pollution gas for a long time, the harm of cultural relics will be very obvious. Therefore, the air quality monitoring system of cultural relics should meet the characteristics of high sensitivity, high reliability, small size, good anti-interference ability, and real-time online detection of multiple gases. The system accuracy test results are shown in Table 3 and Figure 8. It can be seen from the experimental results that the accuracy of the system on the first day is not high enough due to the lack of sufficient sample data in the early stage. However, with the increase of the number of training samples, the accuracy of the system output has also been improved. In addition, the output of the control group has no obvious relationship with the input of the sample number; that is, the increase of the sample number cannot improve the performance of the system without the genetic fuzzy neural network, and the performance is obviously slightly better than that of the complete system. It shows that the intelligent computing module has a good practical effect in predicting user behavior and realizing intelligent control.

There are many exhibition halls in the museum, and the objects that need environmental protection are also different. Therefore, the data of monitoring points are also increasing, and the data tends to show a massive trend. Using traditional data analysis and processing methods to process massive amounts of monitoring data has low efficiency and cannot effectively give information on the pollution level and quantitative evaluation of monitoring points. Using *K*-means cluster analysis and prediction methods, environmental monitoring data can be effectively processed. The traditional method of predicting trends is nothing more than carrying out a large amount of manual data analysis to predict conclusions, and data mining, as a new data-processing technology, can automatically search for information that is valuable for prediction in large databases and can quickly pass the massive data itself. And, draw a conclusion. The most representative one is the Decision Support System (DSS). Data mining can analyze historical sales data to find the market law that can bring the most profit to the enterprise in the future. It can be seen that predicting trends and behavior is discovering predictive knowledge. The showcase security instrument and host computer software constitute a complete system based on the C/S architecture. The showcase security instrument acts as a server to collect monitoring data and process and send them to the upper computer for storage; the host computer acts as a client to send a connection request to the showcase security instrument and deliver working parameters. The display cabinet setting includes the display cabinet quantity setting and display cabinet parameter setting. The former determines the number of display cabinets displayed on the host computer interface, which is convenient for the adjustment of the number of display cabinets in the exhibition hall; the latter determines the display cabinets connected to the host computer, which facilitates the replacement of the security equipment in the display cabinet and enhances the

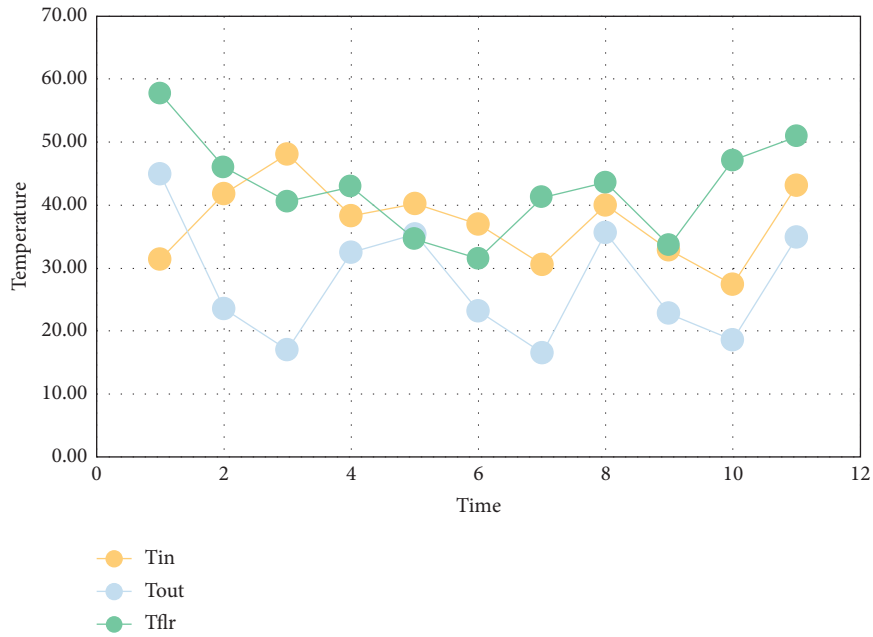


FIGURE 7: Comparison of temperature data.

TABLE 3: System accuracy test results.

| Serial number | 1     | 2     | 3     | 4     | 5     | 6     | 7     |
|---------------|-------|-------|-------|-------|-------|-------|-------|
| Using GFNN    | 0.440 | 0.716 | 0.689 | 0.696 | 0.749 | 0.666 | 0.600 |
| No GFNN       | 0.452 | 0.338 | 0.394 | 0.394 | 0.344 | 0.284 | 0.334 |

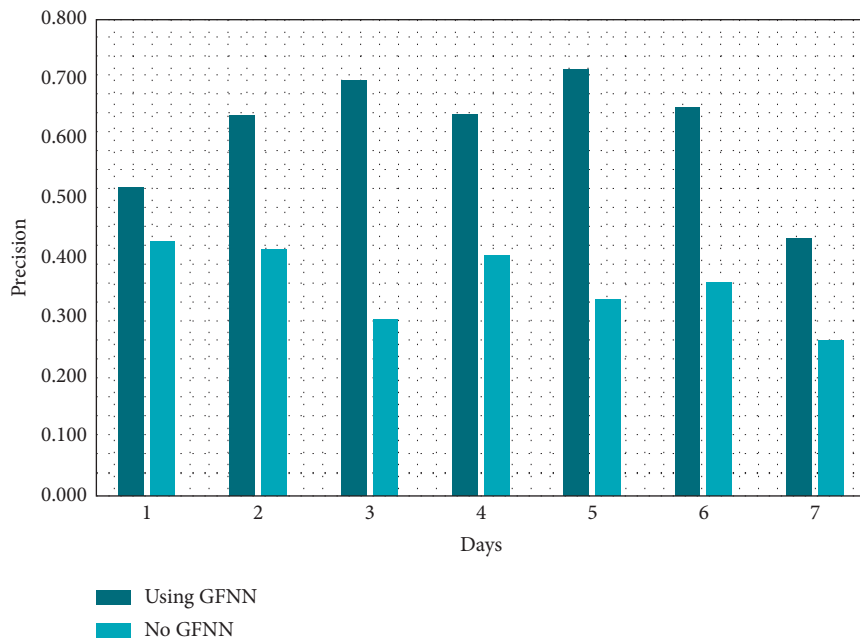


FIGURE 8: System accuracy test results.

intelligence of the museum. The expansibility of the showcase security system makes later maintenance more convenient. The accuracy of the alarm system is shown in

Figure 9. It can be seen from the figure that the application of the Internet of Things technology in the alarm system can significantly improve the accuracy of the system. At the same

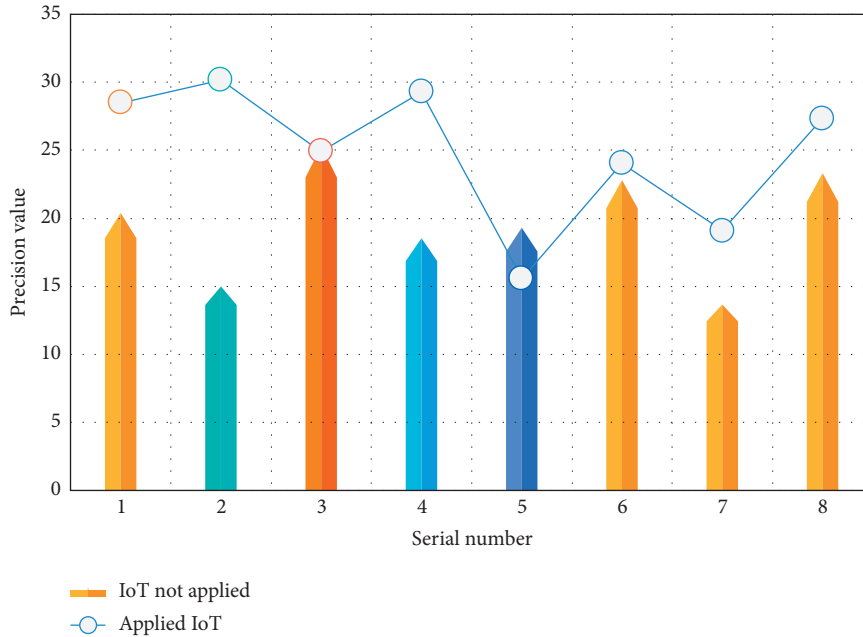


FIGURE 9: The accuracy of the alarm system.

TABLE 4: Comparison results of performance indicators of three control algorithms.

| Controller name            | Overshoot (%) | Steady-state error | Adjustment time (min) |
|----------------------------|---------------|--------------------|-----------------------|
| Conventional PID control   | 36.18         | 0                  | 11.2                  |
| Pure fuzzy control         | 14.34         | 0.003              | 7.4                   |
| Fuzzy adaptive PID control | 8.12          | 0                  | 5.6                   |

time, the sensitivity of the sensor is better, which greatly avoids the destruction of cultural relics.

Compared with the PID control algorithm and pure fuzzy control algorithm, the fuzzy adaptive PID control algorithm has no better control effect in terms of overshoot or other performance indicators. The performance index comparison results of the three control algorithms are shown in Table 4. By comparing the performance indexes of the three control methods, the effectiveness of temperature control in the fuzzy adaptive PID control mode is higher than that of PID control and pure fuzzy control.

## 5. Conclusions

This article mainly studies the research and implementation of the museum's window system based on Internet of Things technology in a smart environment. By introducing the concept, architecture, and basic technologies of the Internet of Things, the article explains the theoretical and practical importance of creating a smart home scene for the Internet of Things. And, with the microcontroller core of the STM32F103xx series as the core, the display casing is monitored through various monitoring channels, such as infrared and smoke. When an alarm occurs, the on-site sound and light alarm shall be sent to the central computer and to the specified mobile terminal at the same time. The radio-controlled method is used to operate the remote control of LED lamps, which can improve the level of

lighting management, save energy, reduce emissions, and reduce maintenance costs.

The system designed in this paper mainly detects the environmental parameters such as temperature, humidity, and illumination in the museum system. The system can effectively detect these data information, so as to achieve the purpose of protecting museum collections. After training the lighting environment data, the dimming classification model is generated, which realizes six levels of intelligent dimming in complex environment and can automatically adjust the dimming level in real time in case of emergency. Through the algorithm, the system realizes the effective energy saving and intelligent control of illumination. At the same time, the hardware circuit of the system is small in size and low in power consumption, which meets the requirements of energy conservation and emission reduction; the software design structure is clear, which is convenient for the later maintenance and upgrading of the system; the test results show that the basic lighting function, Internet of things communication function, and server control function of the system are reliable and meet the design requirements.

With the rapid development of electronic technology, the application scope of the Internet of Things has gradually expanded. From the initial smart home to the intelligent museum showcase system, it is getting closer and closer to our lives. Through the design and construction of the Internet of Things temperature and humidity monitoring system, continuous observation and data collection of the air

temperature, air humidity, and light intensity in the greenhouse are carried out. Through ZigBee communication technology, each node is networked. The control system is equipped with a coordinator, a wireless control node, and multiple wireless detection nodes. The coordinator collects all detection node information and transmits the data to the upper computer through the serial port for display and storage. The monitoring interface is designed through LabVIEW to realize the acquisition, setting, and control of the parameters of each node.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was supported by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

## References

- [1] Z. H. Wu, "Research on the application of internet of things technology to digital museum construction," *Acta Geontica Sinica*, vol. 38, no. 2, pp. 293–298, 2017.
- [2] M. R. Palattella, M. Dohler, L. A. Grieco et al., "Internet of things in the 5G era: enablers, architecture, and business models," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 510–527, 2016.
- [3] O. Bello and S. Zeadally, "Intelligent device-to-device communication in the internet of things," *IEEE Systems Journal*, vol. 10, no. 3, pp. 1172–1182, 2016.
- [4] G. Akpakwu, B. Silva, G. P. Hancke, and A. M. Abu-Mahfouz, "A survey on 5G networks for the internet of things: communication technologies and challenges," *IEEE Access*, vol. 6, no. 12, pp. 3619–3647, 2018.
- [5] S. Liu, "Research on the application of internet of things technology in digital sports and community fitness," *Boletim Tecnico/Technical Bulletin*, vol. 55, no. 4, pp. 139–145, 2017.
- [6] A. V. Dastjerdi and R. Buyya, "Fog computing: helping the internet of things realize its potential," *Computer*, vol. 49, no. 8, pp. 112–116, 2016.
- [7] C. Perera, C. H. Liu, and S. Jayawardena, "The emerging internet of things marketplace from an industrial perspective: a survey," *IEEE Transactions on Emerging Topics in Computing*, vol. 3, no. 4, pp. 585–598, 2017.
- [8] J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang, and W. Zhao, "A survey on internet of things: architecture, enabling technologies, security and privacy, and applications," *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1125–1142, 2017.
- [9] J. Singh, T. Pasquier, J. Bacon, H. Ko, and D. Evers, "Twenty security considerations for cloud-supported internet of things," *IEEE Internet of Things Journal*, vol. 3, no. 3, pp. 269–284, 2017.
- [10] X. Sun and N. Ansari, "EdgeIoT: mobile edge computing for the internet of things," *IEEE Communications Magazine*, vol. 54, no. 12, pp. 22–29, 2016.
- [11] J. W. Xue, X. K. Xu, and F. Zhang, "Big data dynamic compressive sensing system architecture and optimization algorithm for internet of things," *Discrete and Continuous Dynamical Systems-Series S*, vol. 8, no. 6, pp. 1401–1414, 2017.
- [12] I. Yaqoob, E. Ahmed, I. A. T. Hashem et al., "Internet of things architecture: recent advances, taxonomy, requirements, and open challenges," *IEEE Wireless Communications*, vol. 24, no. 3, pp. 10–16, 2017.
- [13] A. Mosenia and N. K. Jha, "A comprehensive study of security of internet-of-things," *IEEE Transactions on Emerging Topics in Computing*, vol. 5, no. 4, pp. 586–602, 2017.
- [14] M. Amadeo, C. Campolo, J. Quevedo et al., "Information-centric networking for the internet of things: challenges and opportunities," *IEEE Network*, vol. 30, no. 2, pp. 92–100, 2016.
- [15] Y. Zhang and J. Wen, "The IoT electric business model: using blockchain technology for the internet of things," *Peer-to-Peer Networking and Applications*, vol. 10, no. 4, pp. 983–994, 2017.
- [16] Z. Longchao, Y. X. Jianjun, and Y. Limei, "Research on congestion elimination method of circuit overload and transmission congestion in the internet of things," *Multimedia Tools and Applications*, vol. 76, no. 17, pp. 18047–18066, 2017.
- [17] H. S. Dhillon, H. Huang, and H. Viswanathan, "Wide-area wireless communication challenges for the internet of things," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 168–174, 2017.
- [18] A. Ouaddah, A. A. Abou Elkalam, and A. Ait Ouahman, "FairAccess: a new blockchain-based access control framework for the internet of things," *Security and Communication Networks*, vol. 9, no. 18, pp. 5943–5964, 2016.
- [19] J. Ni, K. Zhang, X. Lin et al., "Securing fog computing for internet of things applications: challenges and solutions," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 99, pp. 601–628, 2018.
- [20] D. Zhang, L. T. Yang, M. Chen et al., "Real-time locating systems using active RFID for internet of things," *IEEE Systems Journal*, vol. 10, no. 3, pp. 1226–1235, 2017.
- [21] K. Sood, S. Yu, and Y. Xiang, "Software-defined wireless networking opportunities and challenges for internet-of-things: a review," *IEEE Internet of Things Journal*, vol. 3, no. 4, pp. 453–463, 2016.
- [22] S. P. Mohanty, U. Choppali, and E. Kougianos, "Everything you wanted to know about smart cities: the Internet of things is the backbone," *IEEE Consumer Electronics Magazine*, vol. 5, no. 3, pp. 60–70, 2016.
- [23] S. E. Collier, "The emerging enernet: convergence of the smart grid with the internet of things," *IEEE Industry Applications Magazine*, vol. 23, no. 2, pp. 12–16, 2017.
- [24] O. Arias, J. Wurm, K. Hoang et al., "Privacy and security in internet of things and wearable devices," *IEEE Transactions on Multi Scale Computing Systems*, vol. 1, no. 2, pp. 99–109, 2017.
- [25] W. Ejaz, A. M. Naeem, A. Anpalagan, A. Shahid, and M. Jo, "Efficient energy management for the internet of things in smart cities," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 84–91, 2017.
- [26] J. Bernal Bernabe, A. F. Hernandez Ramos, and A. F. Skarmeta Gomez, "TACIoT: multidimensional trust-aware access control system for the internet of things," *Soft Computing*, vol. 20, no. 5, pp. 1763–1779, 2016.
- [27] T. Fossati and H. Tschofenig, "Transport layer security (TLS)/datagram transport layer security (DTLS) profiles for the internet of things," *Physiological Reviews*, vol. 66, no. 4, pp. 1121–1188, 2016.



- [28] Z. Guan, J. Li, L. Wu, Y. Zhang, Y. Wu, and X. Du, "Achieving efficient and secure data acquisition for cloud-supported internet of things in smart grid," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1934–1944, 2017.
- [29] S. Mumtaz, A. Alshaily, Z. Pang, A. Rayyes, K. F. Tsang, and J. Rodriguez, "Massive internet of things for industrial applications: addressing wireless IIoT connectivity challenges and ecosystem fragmentation," *IEEE Industrial Electronics Magazine*, vol. 11, no. 1, pp. 28–33, 2017.

## Research Article

# Application of Remote Sensing Image Data Scene Generation Method in Smart City

**Yuanjin Xu** 

*Institute of Mathematical Geology and Remote Sensing Geology, School of Earth Resources, China University of Geosciences, 388 Lumo Road, Wuhan 430074, China*

Correspondence should be addressed to Yuanjin Xu; xuyj@cug.edu.cn

Received 23 October 2020; Revised 14 December 2020; Accepted 1 January 2021; Published 28 January 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Yuanjin Xu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Remote sensing image simulation is a very effective method to verify the feasibility of sensor devices for ground observation. The key to remote sensing image application is that simultaneous interpreting of remote sensing images can make use of the different characteristics of different data, eliminate the redundancy and contradiction between different sensors, and improve the timeliness and reliability of remote sensing information extraction. The hotspots and difficulties in this direction are based on remote sensing image simulation of 3D scenes on the ground. Therefore, constructing the 3D scene model on the ground rapidly and accurately is the focus of current research. Because different scenes have different radiation characteristics, therefore, when using MATLAB to write a program generated by 3D scenes, 3D scenes must be saved as different text files according to different scene types, and then extension program of the scene is written to solve the defect that the calculation efficiency is not ideal due to the huge amount of data. This paper uses POV ray photon reverse tracking software to simulate the imaging process of remote sensing sensors, coordinate transformation is used to convert a triangle text file to POV ray readable information and input the RGB value of the base color based on the colorimetry principle, and the final 3D scene is visualized. This paper analyzes the thermal radiation characteristics of the scene and proves the rationality of the scene simulation. The experimental results show that introducing the chroma in the visualization of the scene model makes the whole scene have not only fidelity, but also radiation characteristics in shape and color. This is indispensable in existing 3D modeling and visualization studies. Compared with the complex radiation transmission method, using the multiple angle two-dimensional image generated by POV rays to analyze the radiation characteristics of the scene, the result is intuitive and easy to understand.

## 1. Introduction

With the development of science and technology, digital cities have received more and more attention [1]. The concept of digital city originates from the strategic concept of digital earth, also known as network city or smart city, or more precisely information city. It refers to the comprehensive use of computer tools (GIS, remote sensing, telemetry, network, multimedia, and virtual simulation technology). The use of digital technology to collect and process the city's infrastructure and functional mechanisms to enable it to have digital functions, which is conducive to optimizing and improving the city's ecological environment and resources, economy, population, and other complex fields, effectively predicts the future of the city. [2, 3]. The

essence or core of a digital city is the fusion of massive urban spatial data with three-dimensional urban geographic information systems and time series urban geographic information systems [4]. The outstanding feature of digital cities is the ability of applying digital information to grasp the changing process of urban regional structure in time and space. The application research of 3D urban geographic information system and time series urban geographic information system will be an important part of recent digital city theory research. Digital city construction will provide an information security system for the city's sustainable development strategy, meet the government's decision-making, macrocontrol, scientific and technological innovation, natural resources and environmental monitoring, intelligent transportation and urban

management, and various social welfare undertakings, and further provide solutions for the sustainable development of cities. [5, 6]. To build a digital city, we must first apply high-tech means such as computer technology to model the urban environment [7].

Remote sensing image fusion is a process of comprehensive processing the image data obtained by multiple remote sensing sensors or the same kind of sensor for the same target at different times. The image is processed by using certain rules or algorithms, and the useful information contained in the image is fused into a new image. The image contains more accurate and abundant information than any single image, in order to achieve a comprehensive description of the target and ground objects.

The problem of three-dimensional reconstruction of urban buildings has been studied by experts and scholars from various countries for many years and has achieved a series of results. The most representative ones are Google Earth and Microsoft Virtual Earth, which use satellite remote sensing images to generate virtual ground scenes, which have been successfully commercialized on the Internet [8, 9]. Image-based three-dimensional reconstruction of urban buildings is mainly divided into three categories according to the different data sources used: (1) based on remote sensing images, this method uses three-dimensional reconstruction of urban buildings using approximately vertical satellite remote sensing images or aerial images [10]. According to the characteristics of remote sensing imaging, based on the reconstruction of remote sensing images, the reconstruction space is large, and the roof information of the building can also be obtained and the accumulation of errors can be effectively reduced, but the reconstructed buildings have poor fidelity [11]. (2) Based on ground image, ground image-based reconstruction is a three-dimensional reconstruction of urban buildings using images acquired by various ground shooting techniques. According to the characteristics of ground imaging, the reconstruction of this method is better, and the wall texture of the building can be obtained, but the roof information of the building is not obtained, the reconstruction scale is small, and the error accumulation is large [12]. (3) Regarding combination of remote sensing image and ground image, there are advantages and disadvantages in remote sensing and ground-based imaging reconstruction. In fact, remote sensing image and ground image are two important complementary source data. Combining the two for reconstruction is expected to be obtained. This resulted in a reconstruction method combining remote sensing images with ground images. Generally, there are insufficient data acquisition costs, large data volume, complicated calculation, and low automation [13, 14].

Based on the second generation bending wave transform and Dempster-Shafer (DS) evidence theory, Huang C proposed a new remote sensing image fusion method. Huang C uses the bending wave transform to decompose the remote sensing image to obtain the coefficients and uses DS evidence theory to optimize the high coefficients [15, 16]. First, the high-resolution and multiple spectral remote sensing images are decomposed by bending wave transform

to obtain bending wave transform coefficients (coarse, detailed, and fine scale layers) of all layers. Second, the coarse scale layer uses the maximum fusion rule. The detailed scale layer is used by the weighted average fusion rule. The fine scale layer is optimized by DS evidence theory. Three features of the fine scale layer coefficients are obtained. These three characteristics are variance, information entropy, and energy. The use of these features is some parametric belief function and rationality function. The mass function is then combined and a new fusion factor is obtained. Finally, the scene image is obtained by inverse bending wave transform. Rhee et al. attempt to apply two types of image matching, object space based matching techniques and image space based matching techniques, and compare the performance of the two techniques [17, 18]. The object space based match used sets a list of candidate height values for a fixed horizontal position in the object space. For each height, its corresponding image points are calculated and similarity is measured by gray level correlation. The image space based matching used is a modified slack match. Rhee and Kim designed a global optimization scheme for finding the best pair (or group) to apply image matching, defining local matching regions in the image or object space, and merging local point clouds into global point clouds. For optimal pairing selection, the connection points between the images are extracted and a stereoscopic overlay network is defined by using the connection points to form a maximum spanning tree. Qin built the core technology and method related to 3D model reconstruction, focusing on matching of point cloud data registration to simplify denoising, 2D contour extraction, and finally achieving the high complexity of 3D geometric model of farmland site, using advanced 3D printing. Technology produces small 3D printed point cloud data [19, 20].

The innovations of this paper are as follows: (1) introducing the principle of colorimetry method into the visualization of the image simulation scene model and replacing the complex texture with color can reflect the spectral radiation characteristics of the object to a certain extent. Through investigation and research, it is found that the only one that can correspond to the spectral characteristics is its chroma characteristic. Therefore, the chroma is introduced into the scene model visualization so that the whole scene not only is in shape and color, but also has radiation characteristics. This is not available in existing 3D modeling and visualization studies. (2) The remote sensing imaging process can be simulated using the POV ray visual ray tracing software package. It has a convenient and fast programming language, high computational efficiency, and intuitive output. Compared with the complex radiation transfer equation, the multiple angle two-dimensional image generated by POV ray is used to analyze the radiation characteristics of the scene, and the result is intuitive and easy to understand. POV ray can simulate the remote sensing imaging process, mainly because it first defines the position of the light source (sun) and camera (sensor), zenith angle and azimuth, and also sets the camera's field of view, which is another 3D visualization software. And the higher computational efficiency is also relatively advanced in the

field of visualization. (3) Considering the differences of different scenes, the design of remote sensing images for different scenes is different, so that the experimental structure can reflect the diversity and rationality.

## 2. The Proposed Method

### 2.1. Preparation of Remote Sensing Images

*2.1.1. Image Preprocessing.* When the obtained source image is blurred, the contrast is not strong, and the noise interference is large; the corresponding method needs to be used for some processing, so that the subsequent work can be carried out more effectively. Common methods include image enhancement, filtering, histogram correction, and gradation transformation. However, when the quality of the source image is good, it is not necessary to perform these processes. Therefore, the preprocessing of the source image is an optional link, and the processing method is different for different images. The omnidirectional image is used in the experiment in this paper; because the camera has been calibrated before shooting, the lighting conditions are better at the time of shooting, so the obtained image quality is better and generally does not have to be preprocessed.

*2.1.2. Registration.* Registration is to find the mutual correspondence between the omnidirectional map and the remote sensing map and achieve the purpose of reconstruction service through information fusion. Information fusion is the foundation of all subsequent work, so registration is a core component of the entire reconstruction. Conventional methods generally solve the problem of registration between images from homologous images or from the same type of sensor. The omnidirectional and remote sensing images are images formed by heterogeneous sensors. The general reconstruction method of the source image requires registration of the source image, and more registration methods are available at present. However, the existing conventional methods cannot solve the registration problem of remote sensing maps well. The registration of remote sensing maps belongs to the specific key technical problems of 3D reconstruction of remote sensing images.

*2.1.3. High Extraction.* Height extraction is to obtain the height value of the target building in real space, which is an important information of the space structure of the building. The height of the building can be combined with the top view of the building available in the remote sensing map to obtain the approximate spatial structure of the building. Therefore, the height of the building plays a crucial role in the reconstruction of the shape and contour. There are two conventional solutions; one is direct measurement using instruments and equipment, such as laser range finder. The second is to use computer vision principles for estimation. Direct measurement with related instruments is costly. With computer vision estimation, accurate absolute heights are generally not obtained without accurate scale reference objects.

*2.1.4. Shape Modeling.* The goal of shape modeling is to get the outer shape of the entire building. It is also an important part of reconstruction. It is generally modeled based on certain assumptions or prior knowledge (such as a box in the shape of a box and a flat roof) using the obtained height and roof shape information obtained from the remote sensing map. The outer shape of the building is more complex and precise than the approximate spatial structure. The key to shape modeling is to extract the outline of the building. At present, the contour detection algorithms generally have shortcomings such as low detection rate and inability of fully automating. The method of semiautomatic human-computer interaction can be adopted; that is, the existing detection algorithm is combined with the manual correction method for detection. Due to limited time and energy, image-based 3D reconstruction mostly requires shape modeling, and shape modeling methods are almost universal.

*2.2. Building Outline Segmentation and 3D Model Extraction.* How to obtain the singular model of the building from the three-dimensional model of the scene generated by oblique photography is the goal of this paper. DOM can be seen from the high-resolution scene, which has obvious image difference between the building and other features. The image analysis method can be used to extract the outline of the building from the scene DOM and obtain the position information of the building outline, thereby realizing the positioning and segmentation of the building model in the three-dimensional scene model.

*2.2.1. High-Resolution Image E Building Feature Analysis Method.* In an image, the edge information is the most important and basic feature, and the edge feature is the most direct expression and embodiment of the image geometric information. The low-altitude drone oblique photography obtains a higher resolution of the image, and the texture of the scene DOM obtained after the correction is clear. The difference in shape, size, and texture patterns is the basic basis for distinguishing between different features. Through vision, the color of the scene DOM is very realistic, the texture information is very rich, the geometrical structure of the object is more refined, and the recognition of different target objects in the image is more accurate; from a local perspective, a single feature, especially the boundary between the edge of the building and its surrounding environment, is obvious, and the details inside the target object are richly expressed. These features are very advantageous for identifying and extracting the target individual of the building. However, because of the rich information contained in high-resolution images, the phenomenon of "homologous" and "homogeneous foreign matter" appears, resulting in increased noise interference.

In order to extract buildings (houses) in high-resolution images, the characteristics of the buildings are analyzed to establish a good basis for building identification. From spectral and texture characteristics, usually the gray distribution of buildings is relatively uniform. The gray value of the

top is higher than the gray value of other surrounding objects and the texture mode is relatively regular. Generally, the texture performance of buildings has the outline direction of the building that is approximately uniform or orthogonal. For the shape feature, usually the building has geometrically regular edges and corners, and the whole is presented as a regular polygon. From the spatial distribution characteristics, usually the roads in the urban area will be divided into several blocks like chessboards. The buildings are regularly distributed in the block, and the surrounding objects mainly include roads and tree vegetation, so the roads and buildings have strong spatial associations, and usually densely populated buildings will also be arranged regularly and have similar configurations. When considering the geometric features, spectral features, and spatial distribution characteristics of features in high-resolution images, the algorithm is not mature enough and complicated, having low efficiency, which needs further study. According to the research objectives and needs of this paper, the rough positioning of the target of the suspected building does not require accurate and complete extraction of the building. Therefore, this paper only analyzes the geometric features of the building from the geometric shape. In the scene DOM, the geometry of buildings and other features is significantly different. From a straight-down perspective, no matter which building is an individual wrapped by its outer contour, in the high-resolution scene DOM, the outer contour of the building is a connected area with a certain length and area. As a basis for identifying buildings, the internal structure of vegetation such as trees is disorderly, and the boundaries of the whole forest are not clear enough and there are no rules: there are often no clear and regular boundaries on both sides of the road, and the roads are lacking in certain areas. From the perspective of individual buildings, there are also separate topological relationships between different buildings. In the scene DOM, the target features that can be approximated as connected areas tend not to have only buildings, but the buildings also contain other distinguishable geometric features, and the more prominent common features are the length of the outer contour of the building. The outer contour of the building must contain at least a certain number of linear features, according to which the connected areas surrounded by the outline of the building can be further identified and screened.

Therefore, by analyzing the features of the building on the image, the difference and segmentation between the building and other features and buildings can be realized, and the outline of a single suspected building can be extracted and used as a basis for the three-dimensional scene model. A rough monomer model was extracted.

**2.2.2. Building Edge Feature Detection Method Based on Canny Operator.** Edge detection is to obtain information about shape and reflection or transmittance in an image. It is a basic step in image processing, analysis, understanding, pattern recognition, and computer and human vision, being

a very important technology. There are many edge detection methods, such as Roberts operator, Pruitt operator, Sobel operator, and Laplace operator. These are the operators that detect features through local window and are sensitive to noise. Canny proposed the best edge detection operator Canny operator. The operator determines the edge pixels by the maximum value of the image signal function, and the detection performance is good, which has been widely used. Therefore, this paper uses Canny operator to perform edge detection on scene digital image. The scene DOM is a true color image, which needs to be grayed out. The color image can be converted into a grayscale image by using the following formula:

$$P(x, y) = 0.3 * R + 0.59 * G + 0.11 * B, \quad (1)$$

where  $P(x, y)$  is the gray value of the pixel at the  $(x, y)$  coordinate and  $R$ ,  $G$ , and  $B$  are the values in the red, green, and blue primary color channels in the pixel at the  $(x, y)$  coordinate, respectively. There are four main steps in detecting the edge features of grayscale images using the Canny operator:

(1) *Eliminate Noise.* The differential algorithm is highly sensitive to noise, and the Gaussian smoothing filter is used to convolve the image before edge detection to reduce noise interference. The first following formula is a two-dimensional Gaussian function. The principle of Gaussian smoothing is the discrete Gaussian function. The Gaussian function value on the discrete point is used as the weight. For each pixel in the gray image, it is within the window neighborhood of a certain size. Considering pixel weighting to eliminate Gaussian noise, the second following formula is a discrete Gaussian function weight window template with a window size of  $5 \times 5$  pixels, the third following formula is a convolution formula for Gaussian filtering of image  $J$ , and  $g$  is the result of convolution:

$$f(x, y) = \frac{1}{2\pi\sigma^2} e^{-((x^2+y^2)/2\sigma^2)}, \quad (2)$$

$$K = \frac{1}{159} \begin{bmatrix} 2 & 4 & 5 & 4 & 2 \\ 4 & 9 & 12 & 9 & 4 \\ 5 & 12 & 15 & 12 & 5 \\ 4 & 9 & 12 & 9 & 4 \\ 2 & 4 & 5 & 4 & 2 \end{bmatrix}, \quad (3)$$

$$g = I * K. \quad (4)$$

(2) *Calculate the Image Gradient Amplitude Value and the Gradient Direction.* The first-order finite difference is used to approximate the gray gradient of the image. In the Canny operator, the first following formula is used to convolve in  $x$  and  $y$  directions of the image. As shown in the second and third following formulas, the Sobel template is shown,

wherein  $S_x$  is a convolution template in the horizontal direction, and  $S_y$  is a convolution template in the vertical direction:

$$\begin{aligned} G_x &= g^* S_x, \\ G_y &= g^* S_y, \end{aligned} \quad (5)$$

$$S_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \quad (6)$$

$$S_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}. \quad (7)$$

According to the following formulas, the gradient magnitude value  $G$  and the direction  $\theta$  of the image can be, respectively, calculated and the gradient direction is approximated to be generally  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$ :

$$G = \sqrt{G_x^2 + G_y^2}, \quad (8)$$

$$\theta = \arctan\left(\frac{G_y}{G_x}\right). \quad (9)$$

(3) *Nonmaximum Suppression.* The pixel corresponding to the local maximum point is found and retained or marked as an edge pixel, and the gray value of the pixel of the non-maximum point is suppressed and set as the background. This step is mainly to discriminate and remove nonedge pixels, leaving only the candidate image edges.

$$A = \sum \left[ Si^* \log\left(\frac{Si}{Qi}\right) \right]. \quad (10)$$

The image edge is solved using a hysteresis threshold algorithm. The method of lag threshold is to use two thresholds (high and low). The following discriminant basis is used when determining the true edge and removing the false edge: if the gradient amplitude value of a pixel is greater than the high threshold, it is determined that the pixel is a true edge pixel and is retained; if the gradient amplitude value of a pixel is less than the low threshold, it is determined that the pixel is not a true edge pixel and is excluded; if the gradient amplitude value of a pixel is between the high and low thresholds, the pixel has only one gradient amplitude value. When pixels larger than the high threshold are connected, it is determined that the pixel is reserved for the real edge pixel. The scene digital orthophoto is rich in texture, and the geometric features of the features are complex. The edge obtained by the Canny operator for edge detection usually contains a lot of noise. Therefore, the DOM edge detection result needs to be Gaussian. In order to facilitate the identification of the building outline, it is necessary to highlight the contour edge features of the target such as a building from the numerous edge feature pieces of

information in the scene and further suppress the edge pixels whose edge features are not obvious into the background. Therefore, it is necessary to perform Gaussian smoothing on the scene edge detection result and then perform binary processing. Treated by formula (10), we get

$$g_{\text{new}}(x, y) = \begin{cases} 255 & (g_{\text{old}}(x, y) > T), \\ 0 & (g_{\text{old}}(x, y) \leq T), \end{cases} \quad (11)$$

In the formula,  $g_{\text{new}}(x, y)$  is the new gray value of the pixel at the  $(x, y)$  coordinate;  $g_{\text{old}}(x, y)$  is the previous gray value of the pixel;  $T$  is the gray threshold set according to experimental experience and belongs to 0–255. If the gray value of a pixel is greater than the threshold, the gray value of the pixel is set to 255. If it is not greater than the threshold, the gray value is set to 0, and the background is suppressed, thereby obtaining a scene edge with a distinct target edge feature.

### 2.3. Construction of the Original Features of the Front Image.

According to the previous preset, the output of the CNN convolutional layer is more than fully connected:

$$a_k = \sum_{y=1}^H \sum_{x=1}^W w_{x,y} b_{x,y} \cdot f_{x,y,k}. \quad (12)$$

Among them,  $w_{x,y}$  is the response weight at the point  $(x, y)$  in the feature map, and  $b_{x,y}$  is the depth weight of the point. For the response weight matrix  $W \in R^{W \times H}$ , we use the feature map  $f_k$  to construct

$$W = \sum_{k=1}^N f_k. \quad (13)$$

For the depth weight, we use the depth information of the image to assign the weight. We first scale the depth map of the input image to  $W \times H$ ; then,

$$d_{x,y} = \left( \frac{d_{\text{max}} - d(x, y)}{d_{\text{max}} - d_{\text{min}}} \right) + \gamma. \quad (14)$$

Among them,  $d_{\text{max}}$  is the maximum depth,  $d_{\text{min}}$  is the minimum depth, and  $d(x, y)$  is the depth information at  $(x, y)$ .  $\gamma$  is a very small amount to ensure that the depth estimation process of the monocular image is very far away (the misjudgment). The weight value tilts the image feature to the close range.

For the obtained weight matrix, we use the  $L2$  norm to normalize; namely,

$$V = \frac{V}{\|V\|_2}. \quad (15)$$

We perform the above sum pooling calculation on all  $N$  feature maps output by the convolutional layer  $L$  to obtain an  $N$ -dimensional feature vector  $\varphi$  of the convolutional layer, and use the same-dimensional PCA whitening, and then perform the whitening features obtained. The  $L2$  norm is normalized, and finally  $N$ -dimensional image features are obtained.

For comparison, we extract the SPOC algorithm as follows:

$$a_k = \sum_{y=1}^H \sum_{x=1}^W c_{x,y} f_{x,y,k}. \quad (16)$$

Among them,  $c_{x,y}$  is the Gaussian center prior, and its weights are set as follows:

$$c_{x,y} = \exp\left(-\frac{(y - H/2)^2 + (x - W/2)^2}{2\partial^2}\right). \quad (17)$$

Among them,  $\partial$  is the distribution covariance, which is set to one-third of the length of the feature map center from the nearest boundary. It can be seen that the SPOC algorithm adds a Gaussian center prior on the basis of sum pooling and does not effectively reflect the icon in the image. Objects cannot reflect the characteristics of close-up shots.

### 3. Experiments

**3.1. Data Acquisition.** Remote sensing is a means of collecting electromagnetic fields, force fields. Remote sensing maps record this information as an image. The classification of remote sensing maps is more complicated due to differences in sensors, imaging conditions, and types of information collected. They are also treated differently. The proposed method is suitable for visible light imaging and satisfies the remote sensing map of the vertical parallel projection imaging model. Considering the cost and simplicity of acquisition, this paper uses a satellite remote sensing map downloaded from the "satellite" mode of Google Maps. Its resolution accuracy is acceptable, it can be downloaded as long as it can be connected to the Internet, and it is completely free, so it is not only simple and practical but also low in cost. Its data components are more complex, mainly from Digital Globe and MDA Federal. Its imaging feature is a high-altitude bird's-eye view, which provides information on the roof of the building and covers a large area. In addition, it is visible light imaging and approximate vertical shooting; it can be assumed that the downloaded remote sensing image conforms to the vertical parallel projection imaging model; that is, it can meet the requirements of the algorithm.

#### 3.2. Scene Visualization

**3.2.1. Basic Steps of Reverse Tracking.** According to the set image size, the number of rays is determined to be slightly higher than the total number of pixels in the image. If the image size is 160\*120, the total number of pixels is 19200, and the light is 22630; if the image size is 640\*480, the number of pixels is 307200, and the light is 363388; when the image size is 1024\*1280, the number of pixels is 1310720, the light is 1550877, and the number of rays is more than 18% of the number of pixels, so that each cell has a light, and the extra light can be used to verify the correctness of each cell

calculation. When the number of rays is determined, tracking begins. The specific tracking process is as follows:

Step 1: Determine the position of the sensor and the viewing plane, and the light is directed to the scene through the viewing plane;

Step 2: When the light reaches the set opaque surface, the light source is tracked according to the surface reflection principle of the object. At this time, no other object is blocked between the light sources, and the reflection portion is a bright portion;

Step 3: When the light is reflected by the opaque surface, the scene entity that is set when tracking the light source is occluded (the sphere in the figure), and the reflective surface is dark;

Step 4: When the light hits the solid in the scene (the sphere in the figure), it is reflected to the opaque surface, and then the reflection can be traced to the light source. The reflection of the object is bright and there are some optics of the opaque surface characteristic; the light reverse tracking step is shown in Figure 1.

**3.2.2. Acceleration Algorithm.** In the process of ray tracing by POV ray, there are a large number of rays intersecting the object. In order to improve the judgment efficiency of the intersection of light and object, POV ray uses a variety of acceleration algorithms, including multiple layers nested bounding box algorithm. The most important multiple levels nested bounding box algorithm is introduced here, depending on the buffer algorithm and the ray buffer algorithm. The bounding box algorithm is widely used in ray tracing. The traditional bounding box algorithm divides the scene into virtual cubes (such as bounding boxes). First, it is determined which intersecting box the light intersects. If it intersects, it is judged whether the light and the entity in the bounding box are intersecting; compared with the original algorithm that uses image space as an order, this can greatly reduce the number of judgments and improve efficiency. The traditional bounding box algorithm steps are shown in Figure 2.

Since not every bounding box contains entities, POV ray uses a multiple layers nested bounding box algorithm. The multiple levels nested bounding box is similar to the structure of a tree, and the whole scene is first divided into larger bounding boxes. It is judged whether the light intersects with the bounding box. If they intersect, the intersecting bounding box is decomposed into smaller bounding boxes, and the judgment is performed again, and the layers are subdivided into multiple layers in order. In this way, when the entities are discretely distributed in the scene, the computational efficiency can be greatly improved; however, when the entities are continuously distributed, the efficiency is not as good as the traditional bounding box algorithm. Therefore, when using the multiple layers nested bounding box algorithm, the key is to start the bounding box. In POV ray, Bounding = on/off controls whether to use the bounding box, and Bounding Threshold =  $n$  controls the

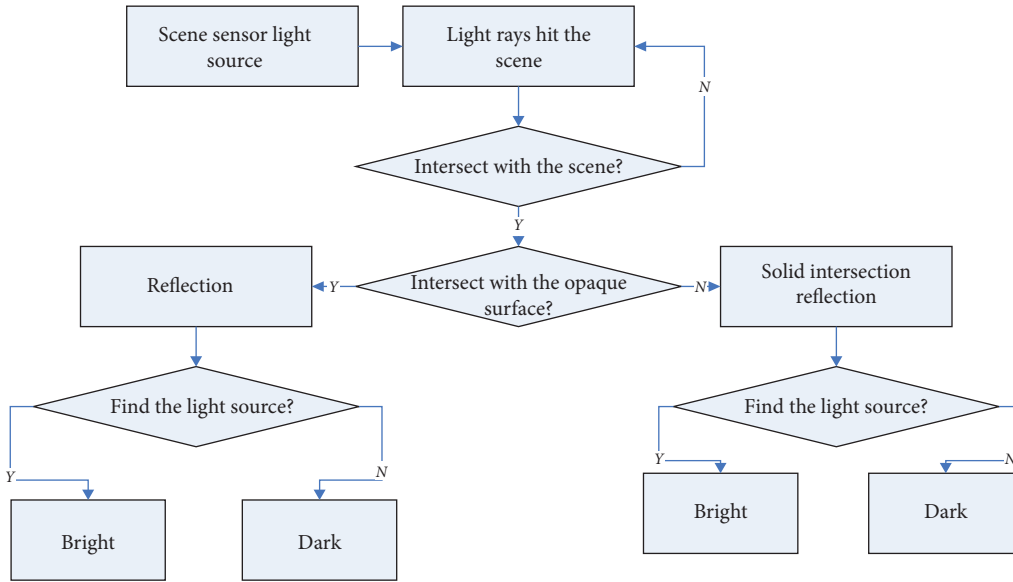


FIGURE 1: Light reverse tracking step.

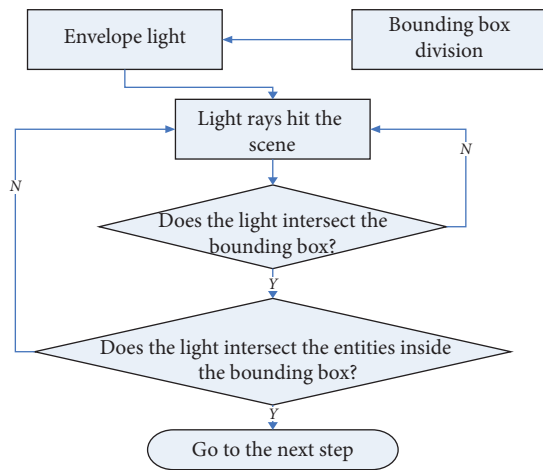


FIGURE 2: Traditional bounding box algorithm steps.

starting layer number  $n$  of the bounding box. POV ray defaults to 3 layers.

**3.3. Generation of House Scenes.** The three-dimensional model of the house can be divided into two parts: the main body of the house and the roof. Therefore, the generation of the house scene can be divided into two parts: the main body of the house and the roof. Considering the complexity of the actual building, this article has been simplified accordingly: the complex building is broken down into a simple four-corner house model. For four-corner houses, there are roughly two categories, one is the most common flat-top houses, and the other is non-flat-top houses. To build a three-dimensional model of a house, you must first obtain some information about the house from some way. Undoubtedly, the information related to the house is mainly the corner coordinates of the house, the elevation of the bottom,

and the height of the house; for nonflat houses, the information about the height of the roof of the house is also known.

**3.3.1. Acquisition of Corner Information of Houses.** The house generally appears as a relatively regular shape in the two-dimensional image, so this paper decomposes it into a rectangle or a square, so that, for the house in the south of the south, it is only necessary to extract the coordinates of the two corners of the diagonal of the house, but for nonpositive south and north houses, the coordinates of the four corners of the house do not have a mutual relationship and must be extracted. In order to be able to generate the above two types of house models at the same time, this paper extracts all the corner coordinates of the house. So, how to extract corner coordinates has become a focus of this article. There are two ways to extract the corner information. One is to extract the boundary information of the house as a straight line segment and then find the intersection point of the straight line segment as the coordinates of the corner of the house; the other is to extract directly according to the gray information of the image (corner coordinates). By comparison, the latter method is found to be simpler and more accurate than the first method. When directly using the image gray scale to extract the angular coordinates of the house, a corner feature extraction algorithm such as a Harris operator and a Moravec operator is often used, which are all difficult to meet the needs of this study to extract the corner coordinates of the house. But with MATLAB's powerful matrix processing capabilities, you can easily solve the problem.

**3.3.2. Acquisition of the Elevation of the Bottom of the House.** Using the classification map, we obtained the coordinate information of the four corner points on the bottom of the house. The acquisition of the elevation of the bottom of the



house is simpler. Since the bottom of the house in the study area is generally flat, the elevation of the bottom surface is uniform. The DEM elevation value of the selected area has been assumed to be 0 before, so the bottom elevation value is 0 from the DEM.

## 4. Discussion

### 4.1. Remote Sensing Image Registration Analysis

*4.1.1. Registration Error.* The registration error is shown in Table 1 and Figure 3. It can be seen from the table that the camera optical axis position error is below 1.3 meters. The result of this accuracy is acceptable. The result has reached the practical standard and is large. In some cases, the resolution of the satellite remote sensing map is also meter level or lower.

The error of the registration algorithm mainly comes from two aspects, one is the error of the experiment itself, and the other is the algorithm error. For the omnidirectional diagram, errors in the relative positional mounting of the components of the camera, errors in the attitude of the device at the time of shooting (such as the level of the device), errors in the imaging device itself, etc. can cause errors in the imaging of the omnidirectional image. Remote sensing maps also have corresponding errors. The error of the experiment itself can be reduced by precisely adjusting the installation of the imaging device and correcting the obtained image. This part will not be discussed in this paper. The registration algorithm error mainly comes from feature extraction, because, from the principle of the algorithm, as long as the extracted features are absolutely accurate, the calculated registration result error is very small. Therefore, the accuracy of feature extraction is the bottleneck of the accuracy of the registration algorithm, especially the principal point registration method, because the features used are basically no information redundancy.

*4.1.2. Registration Time Consuming.* The registration time consumption is shown in Table 2 and Figure 4. It can be seen from the table that both methods are relatively fast and can be controlled within 6 seconds. In comparison, the primary point registration method is faster. Moreover, the time-consuming search registration method will increase with the increase of scene buildings, and the main point registration method will be more time-consuming. This is consistent with theoretical analysis because the equal angle search registration method requires searching for a set of points. And the calculation amount used to judge the feasibility of any point in the point set increases with the increase of the number of linear features used and the principal point registration method only. One or several points need to be calculated, and the amount of calculation for calculating any point is fixed.

### 4.2. Scene Generation Results

*4.2.1. Acquisition of Corners of Houses.* Therefore, based on the classification map obtained from the remote sensing image, the image is read into MATLAB using the `imread` function to generate a matrix, and then the matrix gray information of the house area is given to another matrix of the same size as the original image, so that these two matrices coincide. There are only two values: the gray value of the house and the background gray value. Each house is then marked or given a different color using the `bwlabel` function. Finally, the `regionprops` function can be used to extract the coordinates of the corner points of the region boundary. According to a certain rule, the coordinates of the four corner points of each house can be obtained. The corner points of the house extracted from the classification map are shown in Figure 5.

*4.2.2. Generation of House Scenes.* When obtaining the building model construction information, the various types of information should be stored in the form of a matrix or an array, and then the `patch` function is used to draw the planes. The color of the roof and the color of the surrounding walls are separated, and the storage of the coordinate files is also required. The results of using MATLAB 3D building modeling are shown in Figure 6.

The modeling method of the three-dimensional house is relatively simple, and the required three-dimensional data is small, but the distortion is large, and there is a defect in the aesthetic angle when used for three-dimensional visualization. However, the modeling required in this paper can reflect the radiation characteristics of the ground object. Because the radiation characteristics are only related to the material of the ground under the conditions determined by the lighting conditions and the external environmental conditions, the radiation characteristics of the house are mainly the study of the radiation characteristics of both the exterior wall and the roof of the house. For this purpose, this modeling approach is feasible.

*4.3. Clustering Algorithm on Remote Sensing Images.* Through experiments, it can be concluded that the clustering collective scale of ECUNGA algorithm is set to 5, 10, 20, 30, 40, and 50, the initial random parameter is 20, the maximum allowable algebra of GA is 500, and the maximum allowable stagnant algebra is 50. Perform clustering on the three data sets of Iris, Wine, and Glass, respectively, and compare the best correct rate, average correct rate of 20 times, and worst correct rate of the clustering results under the collective scale of each cluster. The experimental results are shown in Table 3 and shown in Figure 7.

From the experimental results, it can be found that the algorithm is not sensitive to the size of the clustering collective on the data sets Iris and Wine, while the clustering collective scale on the Glass data set is not very stable at 5 and

TABLE 1: Registration error.

|        | Isometric search registration method (m) | Main point registration method (m) |
|--------|--|------------------------------------|
| Test 1 | 1.2955                                   | 0.6411                             |
| Test 2 | 0.4324                                   | 0.6575                             |

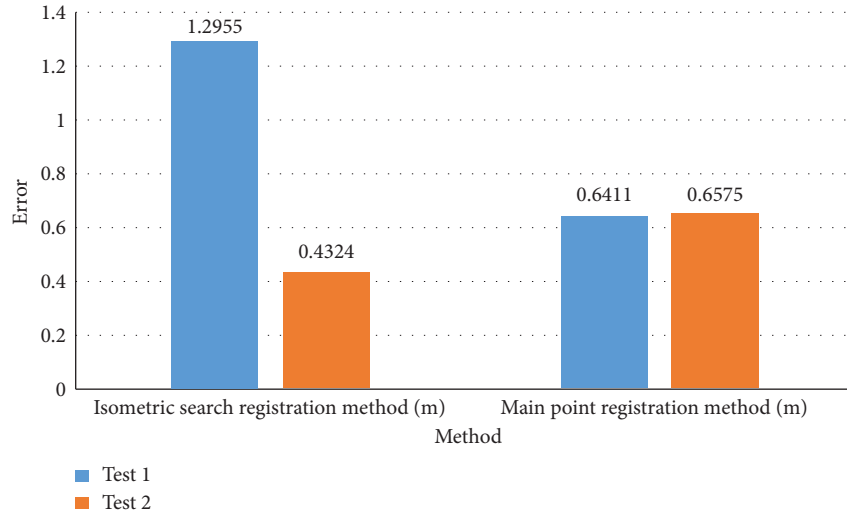


FIGURE 3: Registration error.

TABLE 2: Registration time.

|        | Isometric search registration method (m) | Main point registration method (m) |
|--------|--|------------------------------------|
| Test 1 | 5.058                                    | 0.162                              |
| Test 2 | 1.813                                    | 0.114                              |

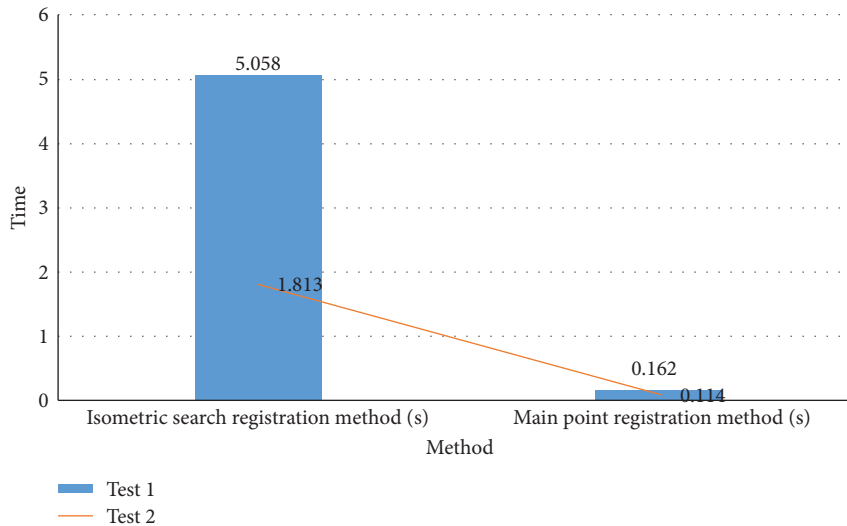


FIGURE 4: Registration time.

10, the best and worst results. The difference is relatively large.

4.4. *Remote Sensing Image Registration.* Analyze the experimental part and get Tables 4 and 5 and Figure 8.

Figure 9 shows the relationship between the number of sampling times of the RANSAC algorithm and the number of last correct matching point pairs and the execution time of the RANSAC algorithm. Red, green, and blue curves represent the experimental results on the test image pairs PA, PB, and PC, and the time unit is seconds. It can be

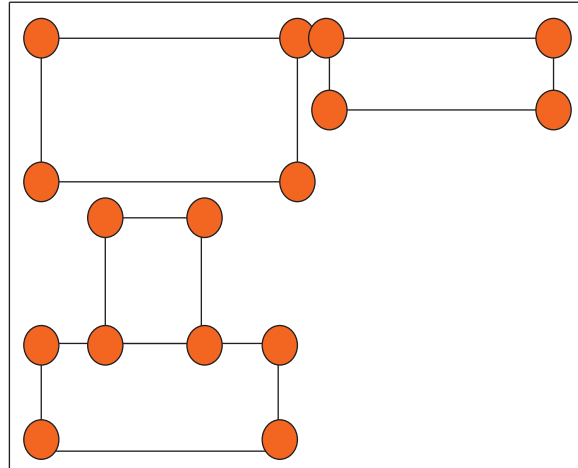


FIGURE 5: House corner extraction effect map.

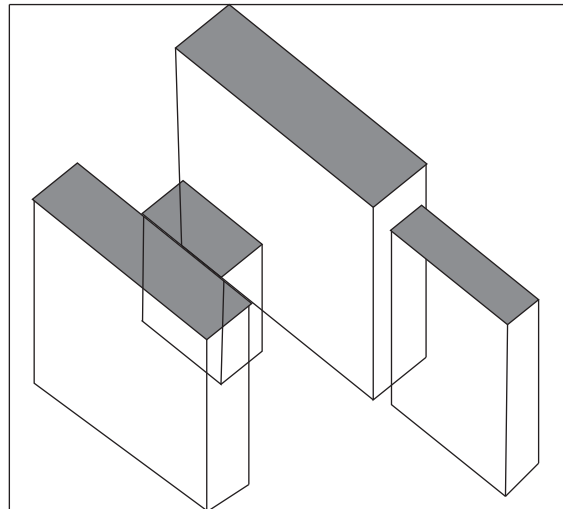


FIGURE 6: Housing scenario generation results.

TABLE 3: The influence of remote sensing image clustering algorithms.

| Data set | Scale | Best | Average | Worst |
|----------|-------|------|---------|-------|
| Iris     | All   | 0.89 | 0.89    | 0.89  |
| Wine     | All   | 0.7  | 0.7     | 0.7   |
|          | 5     | 0.56 | 0.529   | 0.46  |
|          | 10    | 0.54 | 0.48    | 0.46  |
| Glass    | 20    | 0.54 | 0.52    | 0.52  |
|          | 30    | 0.54 | 0.53    | 0.52  |
|          | 40    | 0.54 | 0.51    | 0.52  |
|          | 50    | 0.54 | 0.53    | 0.52  |

clearly seen from the figure that the number of correctly matched feature points remains unchanged after the number of iterative samples is greater than 80, so the threshold value is set to 100 in the RANSAC algorithm based on the similarity transformation model. RANSAC algorithm execution time and sampling times meet a

linear relationship, so the RANSAC algorithm has a higher execution efficiency. It is worth noting that the RANSAC algorithm based on affine transformation usually requires more iterations. Since SIFT is a classic image registration algorithm based on scale space and point features, the SIFT algorithm is still compared here.

TABLE 4: The relationship between algorithm RANSAC sampling times and correct matching point pairs.

| Sampling times | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |
|----------------|----|----|----|----|----|----|----|----|----|-----|
| PA             | 2  | 8  | 8  | 8  | 8  | 8  | 8  | 8  | 8  | 8   |
| PB             | 13 | 20 | 27 | 27 | 27 | 27 | 32 | 32 | 32 | 32  |
| PC             | 76 | 76 | 76 | 76 | 76 | 91 | 91 | 91 | 91 | 91  |

TABLE 5: The relationship between algorithm RANSAC sampling times and running time.

| Sampling times | 10    | 20    | 30    | 40    | 50    | 60    | 70    | 80    | 90    | 100   |
|----------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| PA             | 0.05  | 0.021 | 0.026 | 0.039 | 0.04  | 0.045 | 0.047 | 0.06  | 0.07  | 0.08  |
| PB             | 0.054 | 0.007 | 0.017 | 0.021 | 0.023 | 0.028 | 0.034 | 0.042 | 0.05  | 0.056 |
| PC             | 0.045 | 0.017 | 0.021 | 0.031 | 0.034 | 0.037 | 0.041 | 0.054 | 0.063 | 0.08  |

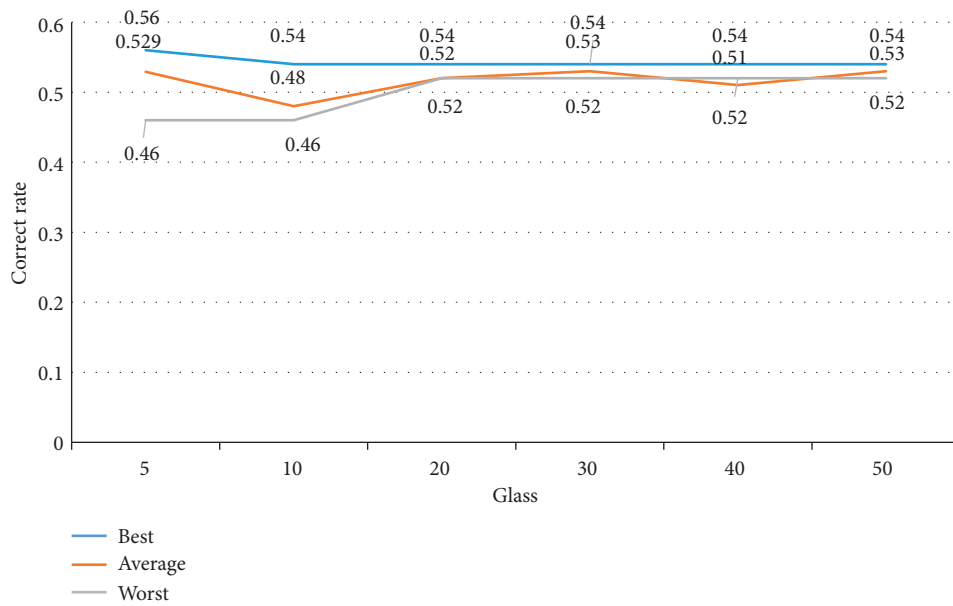


FIGURE 7: The influence of remote sensing image clustering algorithm.

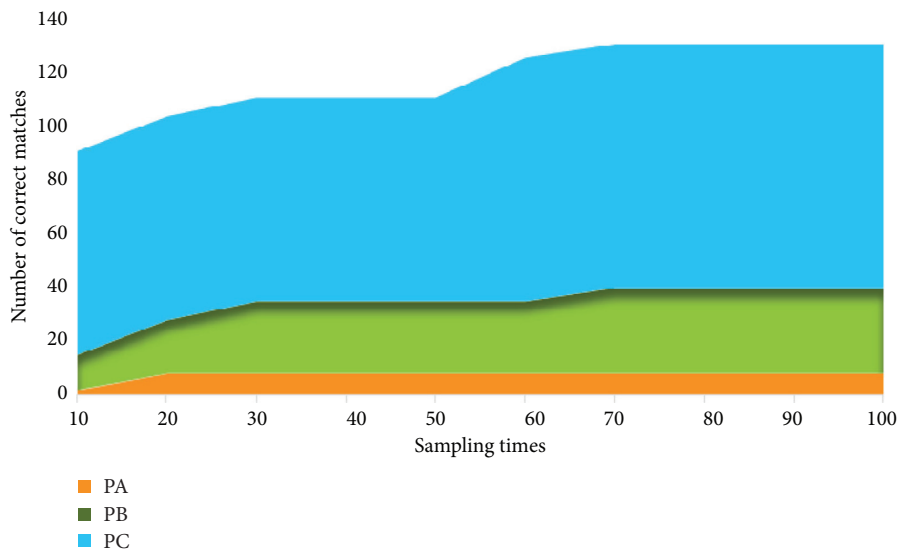


FIGURE 8: The relationship between the number of sampling times of the algorithm RANSAC and the correct matching point pair and running time.

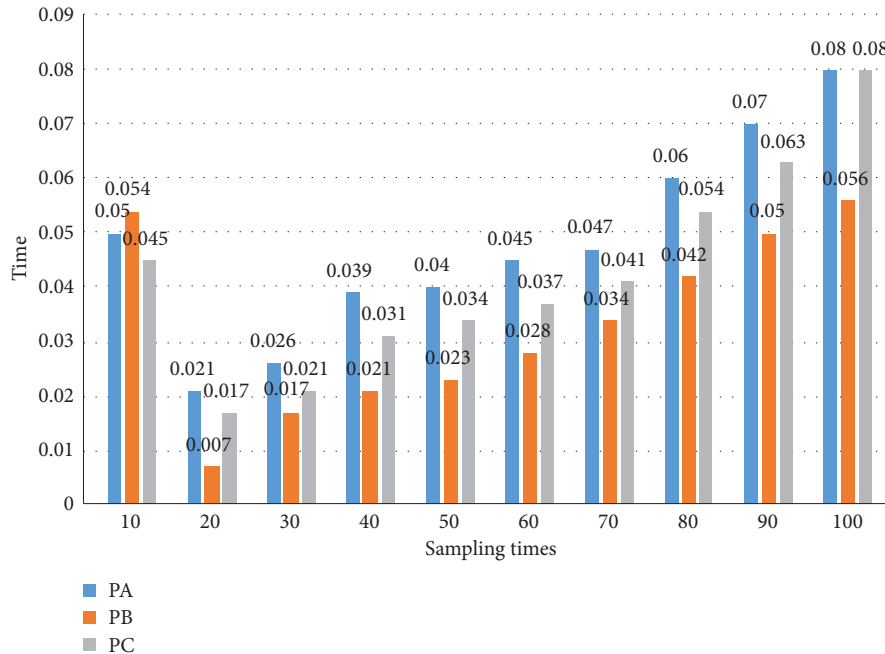


FIGURE 9: The relationship between the number of samples of the algorithm RANSAC and the running time.

It shows the image after registration and fusion using the algorithm proposed in this chapter. It can be seen that the edges and regions overlap well. Therefore, it can be judged intuitively that the registration result is accurate, which again proves the effectiveness of the algorithm proposed in this chapter.

## 5. Conclusions

This paper analyzes and elaborates the concept of hyperspectral remote sensing system and the absorption and reflection of electromagnetic waves. The structure of the hyperspectral scene system is analyzed, and the influencing factors of solar radiation and atmospheric effects and ground reflectivity model are introduced. The imaging mode and imaging principle of the imaging spectrometer were studied and discussed, and the parameters of the remote sensing system of the two scenes were determined.

For the simple scene, the simulation principle and implementation method of spatial correlation and spectral correlation in the ground reflectivity model are studied. The research shows that the spectral and spatial correlation of the features makes the spectral reflection of the pixels belong to different locations of the same type of features. The rate curve fluctuates around its mean reflectance. At the same time, the influence of the atmosphere on solar irradiation intensity and atmospheric transmission coefficient is analyzed. The analysis shows that the worse the visibility of the atmosphere, the smaller the solar irradiation intensity and the atmospheric transmission coefficient.

Based on the remote sensing image, a three-dimensional scene model is constructed. Based on the three-dimensional

scene model and the scene digital orthostatic image, a series of processing is extracted from the geometric and texture features of the scene orthostatic image and the contour features of the building. The minimum rule of the outline of the building is used to outsource the rectangle, and then these outsourcing rectangles are used as the constraint domain to segment the three-dimensional scene model to obtain the coarse single model. On this basis, the triangular and vertical classes of the triangular patches in the three-dimensional space are performed (classification). Calculate the roughness of the regional triangular patches. Based on this, combined with the height of the patch, the triangular patches that are not part of the building features in the coarse single model are purified, and then the adjacent patch growth method is used. The number of feature patches generates a more accurate building unit model and stores a single model from a single file, thus enabling automatic singular modeling of buildings in 3D scenes.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The author declares no conflicts of interest.

## Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities, China University of Geosciences, Wuhan (no. CUGQY1911).

## References

- [1] A. Vanolo, "Smartmentality: the smart city as disciplinary strategy," *Urban Studies*, vol. 51, no. 5, pp. 883–898, 2014.
- [2] P. Neirotti, A. De Marco, A. C. Cagliano, G. Mangano, and F. Scorrano, "Current trends in smart city initiatives: some stylised facts," *Cities*, vol. 38, no. 5, pp. 25–36, 2014.
- [3] L. Sanchez, L. Muñoz, J. A. Galache et al., "SmartSantander: IoT experimentation over a smart city testbed," *Computer Networks*, vol. 61, no. 6, pp. 217–238, 2014.
- [4] J. Jin, J. Gubbi, S. Marusic, and M. Palaniswami, "An information framework for creating a smart city through Internet of things," *IEEE Internet of Things Journal*, vol. 1, no. 2, pp. 112–121, 2014.
- [5] Y. Li, W. Dai, Z. Ming, and M. Qiu, "Privacy protection for preventing data over-collection in smart city," *IEEE Transactions on Computers*, vol. 65, no. 5, pp. 1339–1350, 2016.
- [6] F. Leccese, M. Cagnetti, and D. Trinca, "A smart city application: a fully controlled street lighting isle based on raspberry-pi card, a ZigBee sensor network and WiMAX," *Sensors*, vol. 14, no. 12, pp. 24408–24424, 2014.
- [7] J. Gabrys, "Programming environments: environmentality and citizen sensing in the smart city," *Environment and Planning D: Society and Space*, vol. 32, no. 1, pp. 30–48, 2014.
- [8] K. Czyska, "Application of lidar data and 3D-city models in visual impact simulations of tall buildings," *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 3, no. 7, pp. 1359–1366, 2015.
- [9] R. Heno and L. Chandelier, "3D modeling of buildings: outstanding sites," *IEEE Transactions on Ultrasonics Ferroelectrics & Frequency Control*, vol. 50, no. 11, pp. 1429–1435, 2014.
- [10] X. Zhang, Y. Lv, J. Tian, and Y. Pan, "An integrative approach for solar energy potential estimation through 3D modeling of buildings and trees," *Canadian Journal of Remote Sensing*, vol. 41, no. 2, pp. 126–134, 2015.
- [11] H. Son and C. Kim, "Semantic as-built 3D modeling of structural elements of buildings based on local concavity and convexity," *Advanced Engineering Informatics*, vol. 34, no. 1, pp. 114–124, 2017.
- [12] X. Zhang, L. Yang, and Y. Liu, "3D modeling of urban buildings and trees and its application in building-scale solar energy potential mapping," *Journal of Basic Science & Engineering*, vol. 22, no. 3, pp. 415–425, 2014.
- [13] L. Díaz-Vilariño, K. Khoshelham, J. Martínez-Sánchez, and P. Arias, "3D modeling of building indoor spaces and closed doors from imagery and point clouds," *Sensors*, vol. 15, no. 2, pp. 3491–3512, 2015.
- [14] M. Kedzierski and A. Fryskowska, "Terrestrial and aerial laser scanning data integration using wavelet analysis for the purpose of 3D building modeling," *Sensors*, vol. 14, no. 7, p. 12070, 2014.
- [15] C. Huang and W. Bao, "A remote sensing image fusion algorithm based on the second generation curvelet transform and DS evidence theory," *Journal of the Indian Society of Remote Sensing*, vol. 42, no. 3, pp. 645–650, 2014.
- [16] K. Sugihara and Z.-J. Shen, "Automatic generation of 3D building models by rectification of building polygons," *Advanced Science Letters*, vol. 21, no. 12, pp. 3649–3654, 2015.
- [17] S. Rhee and T. Kim, "Dense 3d point cloud generation from uav images from image matching and global optimization," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLI-B1, no. 1, pp. 1005–1009, 2016.
- [18] S. G. Nebaba and A. A. Zakharova, "An algorithm for building deformable 3D human face models and justification of its applicability for recognition systems," *SPIIRAS Proceedings*, vol. 3, no. 52, pp. 157–179, 2017.
- [19] J. Qin, C. Fang, Y. Wang, G. Li, and S. Wang, "Evaluation of three-dimensional urban expansion: a case study of Yangzhou City, Jiangsu Province, China," *Chinese Geographical Science*, vol. 25, no. 2, pp. 224–236, 2015.
- [20] L. Ding, Y. Zhou, and B. Akinci, "Building Information Modeling (BIM) application framework: the process of expanding from 3D to computable nD," *Automation in Construction*, vol. 46, no. 6, pp. 82–93, 2014.

## Research Article

# Modeling and Research on Human Capital Accumulation Complex System of High-Tech Enterprises Based on Big Data

Yanan Shen <sup>1,2</sup>

<sup>1</sup>College of Economics and Management, Harbin University of Science and Technology, Harbin 150080, Heilongjiang, China

<sup>2</sup>Harbin University of Commerce, Harbin 150028, Heilongjiang, China

Correspondence should be addressed to Yanan Shen; shenyn315@163.com

Received 23 October 2020; Revised 28 December 2020; Accepted 13 January 2021; Published 28 January 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Yanan Shen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

At present, high-tech enterprises are mainly organizations engaged in the production, research, and development and service of high-tech products. The current development of high-tech industries in various countries in the world is of great significance to improving social productivity and overall national strength. This article mainly introduces the modeling and analysis of the complex system of human capital accumulation in high-tech enterprises based on big data. This paper proposes a theoretical analysis of corporate human capital data and proposes regression analysis and analytic hierarchy process to calculate and analyze the process of corporate human capital accumulation and collect and analyze the data using questionnaire surveys and other methods. In addition, complex systems were modeled based on big data, and a cloud computing-based human capital big data analysis platform was established. The experimental results show that complex system modeling based on big data has a positive impact on the human capital accumulation of high-tech enterprises. For high-tech companies at different stages of development, the relationship between various types of human capital and corporate performance is inconsistent with the market economy. In contrast, the current salary system of some high-tech companies still has many drawbacks. The cost of using human resources controlled by the wage system accounts for more than 24%.

## 1. Introduction

With the development and dissemination of information technology and computers, the flow of personnel, capital, goods, and information in the production and living activities of human society gradually tends to be digital. Digital products are not uncommon in life, which in turn activates industry data. Explosive development has promoted the rapid development of big data. However, traditional human resource management often lags behind, even simple manual and subjective personnel management. Even the performance appraisal standards are subjective, which run counter to the development of society.

With the emergence of network and information, especially the widespread popularization of computer and Internet technology, we should make full use of the convenience provided by information technology to make human resource management more scientific, reduce

human capital, and give full play to the role of human resource management [1]. Through the production of a large amount of internal and external data, the processing and analysis of a large amount of human capital accumulation data of high-tech enterprises can bring huge added value to enterprises and adapt to the changing economic environment [2]. Through the analysis of the current situation of human resource information system, we found its shortcomings and further put forward the theory and method of creating a new platform, which provides support for the establishment of analysis platform based on high-tech enterprise data of computing cloud [3].

H. Chen, B. Xuan, P. Yang, and others stated that safety requirements are becoming more and more important for various industrial production processes. Although many facilities have been designed and used to ensure safety, some safety accidents still occur [4]. But he did not come up with a plan on how to prevent brain drain. D. Sledgianowski et al.

responded to the initiative to apply big data technology to accounting courses and used the lens of the accounting education ability integration framework to provide examples of integrating big data and information systems into teaching resources [5]. From the accounting of R&D expenses of enterprises, the expenses in the process of R&D are calculated, and reasonable cost sharing standards are formulated. The main purpose of this paper is to improve the cost control of R&D projects of high-tech enterprises, so as to ensure the economic benefits of high-tech enterprises. However, his research on high-tech enterprises is more invested in enterprise benefits and does not consider how to introduce high-tech talents [6]. Z. Wang, J. Yang and others took high-tech companies listed on the small board from 2007 to 2010 as a sample and used the structural equation model to explore the political connection mechanism of technological innovation in small and medium-sized high-tech companies. They said that politically connected managers can play an active role in the technological innovation of small- and medium-sized high-tech companies. External financing and entry barriers play an indirect role in the transmission of political connections to technological innovation, that is, political connections indirectly affect the technological innovation of enterprises through external financing and entry barriers [7]. This strategy is based on fully realizing that the international competition at present and in the future is mainly scientific and technological competition and also a correct decision based on China's national conditions and the rise and challenge of knowledge economy in the world. The enlightenment of human capital investment theory to us: Talent is capital [8]. This view sees the importance of talent but it does not have a good explanation for big data. To attract talents, we should first know where the talents are and rely on big data technology for precise positioning [9]. Although the theory of human capital has been attached importance to by various disciplines, scholars have been studying it more and more. In recent years, most of the researches at home and abroad have focused on the relationship between human capital and economic growth or enterprise performance, or the impact of the combination of human capital and technological innovation on economic growth [10]. The research on the relationship between human capital and technological innovation is scattered, and most of them only use qualitative analysis to analyze the relationship between human capital and technological innovation ability. The mechanism of human capital on technological innovation capability is ignored [11].

This article mainly introduces the modeling and analysis of the complex system of human capital accumulation in high-tech enterprises based on big data. This article puts forward the theoretical analysis of enterprise human capital data, questionnaire survey, and other methods and proposes regression analysis and analytic hierarchy process to calculate and analyze the process of enterprise human capital accumulation. In addition, complex systems were modeled based on big data, and a cloud computing-based human capital big data analysis platform was established.

## 2. Human Capital Acquisition Method of High-Tech Enterprises Based on Big Data Technology

### 2.1. Method of Analyzing the Data of Human Capital of High and New Enterprises

*2.1.1. Methods of Theoretical Analysis.* This paper expounds the background and content of the theory of enterprise capability in detail and expounds the mechanism of the relationship among human capital investment, enterprise capability, and enterprise performance, so as to provide a reasonable theoretical basis and suitable and meaningful hypothesis to be tested for the empirical research in the following paper. Analyze the connotation and characteristics of core human capital and general human capital and the different characteristics between high-tech enterprises and traditional enterprises and understand the commonness and individuality between them [12]. So we have a deeper understanding of the core human capital and its internal relationship with the competitiveness of high-tech enterprises. In the competition of the enterprise, only know the enemy and the friend can win the victory [13]. The theoretical analysis method can help the enterprise to understand the strengths and weaknesses and can analyze the strengths and weaknesses of the competitors [14].

*2.1.2. Questionnaire Analysis.* In this paper, a questionnaire about the relationship between the optimal allocation of human resources and the performance of high-tech enterprises is developed, and some enterprises of high-tech enterprises are surveyed by questionnaire [15]. It is a kind of survey method to collect research materials indirectly in written form, a method of obtaining materials and information indirectly by sending a concise inquiry form to the investigator and filling in the opinions and suggestions on the relevant issues [16]. In this way, we can clearly know the human resources structure of high-tech enterprises, and the following investigation and analysis can be carried out more smoothly [17].

### 2.2. Base on Human Capital Law of High-Tech Enterprises Based on Big Data

*2.2.1. Visual Research Method.* Visual analysis refers to the mutual analysis of a large amount of data, which is affected by the information dispersion and nonuniform data structure [18]. In addition, manual analysis occupies the core means, and the analysis process lacks structure and stability. This leads to the analysis process and mode cannot be clear, which makes it more difficult for the application system to obtain data, which hinders the subsequent mining and processing analysis [19]. The construction of visual data analysis platform effectively solves this problem and realizes association analysis manually. The essence of data visualization is a research system based on service-oriented guidance, which is comprehensive and unique [20]. This theoretical concept first appeared in the 15th century and has not been well developed before the 20th century. With the emergence and update of electronic computers, this



guiding ideology has been reborn [21]. With the continuous promotion and application of “big data,” the efficient graphic carrier of visual analysis makes the analysis and understanding of nonsimple data more concise and can effectively assist the application, prediction, and decision-making of data information [22].

**2.2.2. Predictive Analysis.** Analysis and prediction are the ultimate goal of big data application, which must be based on effective and reliable data sources [23]. The basic data needed by HR management is mainly divided into two parts, internal and external, including market environment factors [24]. Generally, the internal data has good accuracy, but it lacks conciseness in collection and application. In addition to the support of data sources, professional analysis software and big data professional analysts are more efficient predictive analysis methods [25].

**2.3. Principal Agent Theory Based on Utility Function.** The principal-agent theory is completely based on the hypothesis of “economic man.” For a rational economic man, the choice of his working style and effort is based on the maximum utility. The general utility function is as follows:

$$U = U(w, e). \quad (1)$$

Among them,  $w$  refers to employee’s income or remuneration and  $e$  refers to employee’s effort. Generally speaking, income is directly proportional to utility, and effort is inversely proportional to utility. Thus, the following formula is established:

$$\frac{\partial U}{\partial w} > 0. \quad (2)$$

Obviously, employees choose as much as possible in exchange for more remuneration with less effort so as to have

$$\frac{\partial U}{\partial e} < 0. \quad (3)$$

The output of an enterprise is a function of effort and environmental factors:

$$q = q(e, s). \quad (4)$$

In the formula, the environmental factor  $s$  is a random variable, which has an important impact on the output of enterprises.

**2.4. Regression Analysis of Human Capital Calculation.** In the regression analysis, firstly, the two factors extracted are taken as the explanatory variables and the explanatory variables to establish a linear regression equation to analyze in detail the impact of large technology on the human capital of high-tech enterprises. The formula of the regression model

$$Y = C_1X_1 + C_2X_2 + \text{Control} + e_1. \quad (5)$$

The regression model is as follows:

$$\begin{aligned} M_1 &= C_1X_1 + C_2X_2 + \text{Control} + e_2, \\ M_1 &= A_{11}X_1 + A_{22}X_2 + \text{Control} + e_2. \end{aligned} \quad (6)$$

The human capital investment of high-tech enterprises can mainly positively affect the production efficiency and operation ability of the enterprise, but the influence on the management efficiency of the enterprise is not significant and not positive. It focuses on the mediating role of the two dimensions of the ability of the enterprise’s production efficiency and operation ability in the impact of human capital investment on enterprise performance. The formula of specific regression models is

$$Y = C_1X_1 + C_2X_2 + B_1M_1 + B_2M_2 + \text{Control} + e_3. \quad (7)$$

**2.5. Analytic Hierarchy Process of Human Capital.** The AHP is to solve a decision-making problem, which will be solved as a subsystem, and then solve the decision-making problem by analyzing the influencing factors and divide a complex problem into simple problems related to it. In this subsystem, the decision problem to be solved is taken as the goal, and the factors affecting the target level are taken as the standard and then further decomposed at different levels as the standard.

According to the scale value and the method of pairwise comparison, the formula is as follows:

$$\begin{aligned} a_{ij} &> 0, \\ a_{ij} &= \frac{1}{a_{ji}}, \\ a_{ij} &= 1. \end{aligned} \quad (8)$$

## 2.6. Mode of High-Tech Enterprise Entrepreneurs’ Human Capital Participating in Income Distribution

**2.6.1. Determination of Withdrawal Amount of Virtual Stock Option Award Fund.** Before the implementation of the option incentive plan, the general meeting of shareholders shall discuss and determine a basic withdrawal proportion and withdraw a certain amount from the current year’s EVA value as the company’s first period incentive start-up fund, including  $F_0$ , the initial incentive plan start-up fund,  $z_0$ , the enterprise’s EVA value when implementing the option plan for the first time, and  $F_0$ , the withdrawal proportion of the first incentive fund:

$$F_0 = z_0 * f_0. \quad (9)$$

After that, the incentive fund FN of that year will be extracted from the EVA value created in that year according to a certain proportion, so as to reward the virtual stock option. The reward level is linked with the company’s operating performance level and is calculated according to the company’s operating performance according to the specific calculation formula, where FN is the reward fund of the

company in the  $n$ th year;  $Z_n$  is EVA value of the company in the  $n$ th year;  $F_n$  is the proportion of the bonus fund withdrawn in the  $N-1$  year of the company;  $G_n$  is determined by the excess EVA growth rate of the company in the  $n$ th year:

$$\begin{aligned} F_n &= Z_n F_n - 1(1 + G_n) \quad (Z_n > 0 \text{ and } 1 + G_n > 0), \\ F_n &= 0 \quad (Z_n < 0 \text{ and } 1 + G_n < 0). \end{aligned} \quad (10)$$

The calculation method of the excess EVA growth rate is as follows, where the expected value is recorded as E1:

$$G_n = \frac{(\Delta \text{EVA} - \text{EI})}{\text{EVA}_{N-1}}. \quad (11)$$

Since the values of the source data are both negative numbers, all source data need to be taken as absolute values:

$$\begin{aligned} U &= \{P_1|D, L, f_2, Q, d, l \quad P_2|f_1, \mu \quad P_3|N, M, I\}, \\ \text{IR} &= \sum_{s=1}^U \sum_{d=1}^K f_s, DV_s, d. \end{aligned} \quad (12)$$

Take the first 4 values of the RSRP value of the service station received at time  $t$ , and form the short sequence of the GM (1,1) model with the predicted value obtained at the previous time, denoted as

$$\sigma_{ikjl} = \begin{cases} \frac{n}{\Delta_{ikjl}} \sqrt{\sum_{s=1}^n (x_{ik}(\varepsilon) - x_{jl}(\varepsilon))^2 \Delta_{ikjl}(\varepsilon)}, & \Delta_{ikjl} > 0; \\ 0, & \Delta_{ikjl} < 0, \end{cases}$$

$$X_1 = \{X_1(t - 4\Delta t), X_1(t - 3\Delta t), X_1(t - 2\Delta t), X_1(t - \Delta t)\}. \quad (13)$$

In the same way, remember the short sequence formed by the RSRP value of the target site as

$$X_2 = \{X_2(t - 4\Delta t), X_2(t - 3\Delta t), X_2(t - 2\Delta t), X_2(t - \Delta t)\}$$

$$\text{BMI}(b) = 2n \ln(\sigma) + n \ln(2\pi) + n \left\{ \frac{n + tr(S)}{n - 2 - tr(S)} \right\},$$

$$W(T) = K(y(T-1), u(T-d-1)). \quad (14)$$

Establish a GM(1,1) model based on the short sequence  $X$ , and get a new set of prediction sequences after performing operations:

$$\begin{aligned} Y_1 &= \{Y_{1,1}, Y_{1,2}, Y_{1,3}, Y_{1,4}, Y_{1,5}\}, \\ y_i &= \beta(u_i, v_i) + \sum_{j=1}^p \beta_j(u_i, v_i) x_{ij} + \varepsilon_j \beta_j. \end{aligned} \quad (15)$$

In the same way, a GM(1,1) model is established based on the short sequence  $2X$ , and a new set of prediction sequences are obtained after calculation:

$$\begin{aligned} Y_2 &= \{Y_{2,1}, Y_{2,2}, Y_{2,3}, Y_{2,4}, Y_{2,5}\}, \\ w_G^{A_i A_j} &= \max\left\{0, W_G \cdot \varepsilon \left( f_G^{A_i}, f_G^{A_j} \right)\right\}. \end{aligned} \quad (16)$$

Take the average of the predicted sequence of the service site and the target site obtained after the prediction to obtain the average sequence:

$$\begin{aligned} Y'_1 &= \text{avg}(Y_1), \\ Y'_2 &= \text{avg}(Y_2). \end{aligned} \quad (17)$$

### 3. Human Capital Accumulation System Experiment under Big Data Technology

**3.1. Basic Characteristics of Big Data Machine Learning.** The large system of machine data learning includes many complex technical problems in mechanical learning and big data processing, such as model structure, teaching algorithm, accuracy of mechanical learning model, distributed storage, and parallelism of big data processing. These are closely related to calculations, network communications, local calculations, work plans, system fault tolerance, and other factors. However, data development method is a nonparametric estimation method, which uses mathematical programming to evaluate the relative effectiveness of multiple input and output decision-making units, that is, to judge whether DMU is within the "limit" of total production potential. Since production boundary is the extension of production function in the case of multiple costs in finance, DEA method and model can determine its structure. For production boundary, DEA method can be regarded as a nonparametric statistical method. When you use data envelopment analysis to evaluate the effectiveness of decision-making units, you can not only get the information decision-making of relative performance and scale economy of each receiving unit but also obtain a lot of management information with profound economic significance and background.

**3.2. Analytic Hierarchy Process Test.** Analytic hierarchy process (AHP) is a kind of multiple criteria decision-making process that combines qualitative and quantitative analyses. Its characteristic is to establish a hierarchical structure model by analyzing the essence, influencing factors, and internal relations of complex problems and then determine the decision-making process mathematically with less quantitative information, thus providing a simple decision-making method for the resolution of composite materials and providing quantitative basis for analysis, decision-making, prediction, or mathematical control. The following is how to use it:

- (1) According to different objectives and functions, the system is divided into different levels, such as target level, standard level, and system level.

- (2) In pairwise comparison, the detailed level procedure of decision table (AHP) is mainly for people to evaluate the relative importance of each factor at each level. These decisions are expressed in the form of numbers by introducing appropriate scales, reflecting people's understanding of the relative importance of various factors. Generally, the 1–9 scale method and the reciprocal scale method are used. By pairing and comparing the factors, the values are assigned according to the importance level, thus completing the transformation from qualitative analysis to quantitative analysis.
- (3) Hierarchical single sort. It is to calculate the weight of the elements related to the upper level according to the decision table. Theoretically, the calculation of the single level classification of the problem may be limited to the calculation of eigenvalues and the maximum value of the sector of the crisis matrix.
- (4) In order to determine the relative importance or relative advantages and disadvantages of the lower level elements of the overall goal, the comprehensive weight of each element in each layer is calculated and the general classification is carried out.

#### 4. Human Capital Analysis of High-Tech Enterprises Based on Big Data

**4.1. Item Data Analysis Based on Scale.** As shown in Figure 1, the test of item analysis is to explore the difference of subjects with high and low scores in each item or to test the homogeneity among items. The difference between the first 27% and the last 27% of the total score of the scale is called the comparison of the two extreme groups, and the difference value of the comparison results of the extreme group is called the decision value or critical ratio. Data analysis is shown in Table 1 and Figure 2.

As shown in Table 2, *t*-test of independent samples is to test whether the average of each item in the high score group and the low score group is significant, so as to know whether the average of each item in the scale is different due to the different groups (high group and low group). In terms of the variable C1, the average of the high group was 2.5, the standard deviation was 1.4, the average of the low group was 3.7, and the standard deviation was 1.2. The greater the difference between the two groups was, the more likely the difference would be.

**4.2. Factor Analysis of Human Capital.** KMO and Bartlett's test is the fitting test before factor analysis. KMO is an index for comparing the observed correlation coefficient and partial correlation coefficient. The value of KMO reflects the correlation degree of this factor analysis. In general,  $KMO > 1$  is very suitable for factor analysis;  $1 < KMO < 1.1$  is suitable; above 0.9 is acceptable; 0.7 is poor; below 0.4 is not suitable for factor analysis. The results of KMO and Bartlett's test on 12 variables in human capital measurement table are as follows:  $KMO = 0.8 > 0.7569$ , and the significance probability sig of  $\chi^2$  statistical value of Bartlett's sphericity test is

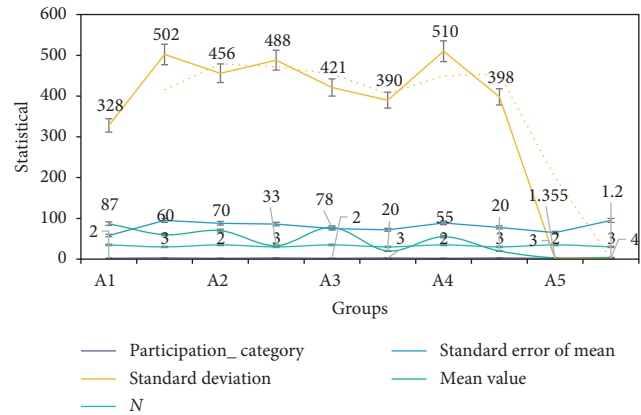


FIGURE 1: Statistical quantity of groups statistical graph.

less than 0.11, indicating that the factors in the questionnaire have certain relevance, as shown in Table 3.

Factor analysis is to study the internal dependence among many variables. The principal component method is used to extract common factors, and the orthogonal rotation method of varimax is used to obtain the load value of each factor. The main purpose is to condense many observation variables into a few factors to achieve the effect of reduction to make the scale more scientific and simplify the later calculation.

**4.3. Data Analysis of Regression Method.** When establishing the regression model of economic problems, it is necessary to consider whether there are conditions inconsistent with the assumptions of the model, such as heteroscedasticity. If the problem of heteroscedasticity exists in the model, the least-square method is still used to estimate the unknown parameters, which will lead to the results that the estimated values are not optimal. The residual graph analysis method of SPSS software is used to test the heteroscedasticity. The test results show that the model does not have heteroscedasticity problem.

As shown in Figure 3, high-tech enterprises' nonexecutive human capital investment and executive human capital investment can positively and significantly affect their performance, and this impact has a certain lag effect. In the control variables, the human capital stock of enterprises can promote the performance of enterprises and is significant, which also confirms that the reasonable human capital structure in high-tech enterprises is the premise to improve the efficiency of human capital investment, and the higher the quality of human capital, the better the performance of enterprises.

As shown in Table 4, when the flow of human capital is more frequent, it will reduce the performance of the enterprise. This may be because when the flow of human capital is more frequent, the recruitment cost that the enterprise needs to invest in the next year will increase, and the newly introduced human capital often has a certain running in period with the enterprise, which will have an adverse impact on the performance of the enterprise.

TABLE 1: Statistical quantity of groups.

| Participation_category | N | Mean value | Standard deviation | Standard error of mean |
|------------------------|---|------------|--------------------|------------------------|
| A1                     | 2 | 35         | 87                 | 328                    |
| A1                     | 3 | 30         | 60                 | 502                    |
| A2                     | 2 | 35         | 70                 | 456                    |
| A2                     | 3 | 30         | 33                 | 488                    |
| A3                     | 2 | 35         | 78                 | 421                    |
| A3                     | 3 | 30         | 20                 | 390                    |
| A4                     | 2 | 35         | 55                 | 510                    |
| A4                     | 3 | 30         | 20                 | 398                    |
| A5                     | 2 | 35         | 3                  | 1.355                  |
| A5                     | 3 | 30         | 4                  | 1.2                    |

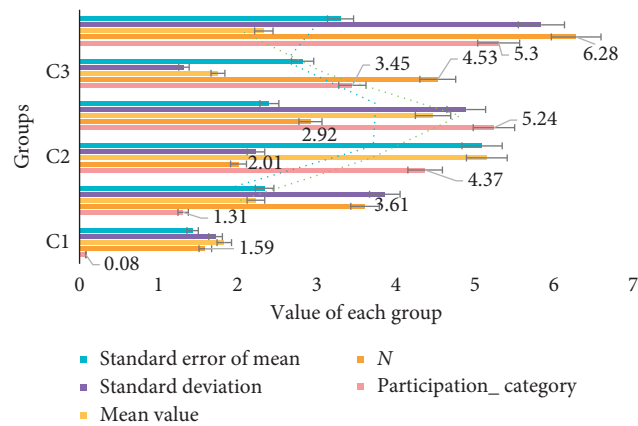


FIGURE 2: The average value of the high and low groups.

TABLE 2: Comparison of the standard deviation of the average value of the high and low groups.

| Participation_category | N | Mean value | Standard deviation | Standard error of mean |
|------------------------|---|------------|--------------------|------------------------|
| C1                     | 2 | 35         | 3.15               | 1.366                  |
| C1                     | 3 | 30         | 4.72               | 1.35                   |
| C2                     | 2 | 35         | 2.5                | 1.225                  |
| C2                     | 3 | 30         | 3.9                | 1.088                  |
| C3                     | 2 | 35         | 1.235              | 1.365                  |
| C3                     | 3 | 30         | 1.265              | 1.965                  |

TABLE 3: KMO and Bartlett's test.

|   |        |
|---|--------|
| Kaiser-Meyer-Olkin measure of sampling adequacy | 800    |
| Approximate chi square                          | 605.29 |
| Bartlett's sphericity test                      | 88     |
| Sig   | 1      |

4.4. HR Staff's Views on Big Data. The most important part of human resource planning is to investigate the current human resource situation of the company in the preparatory stage. The quantity and quality of all human resources of the company are the foundation of enterprise development. In general, the number of existing employees, the number of people of all ages, the distribution of employees at all levels of education, and the skills of employees are the contents of the company's thinking.

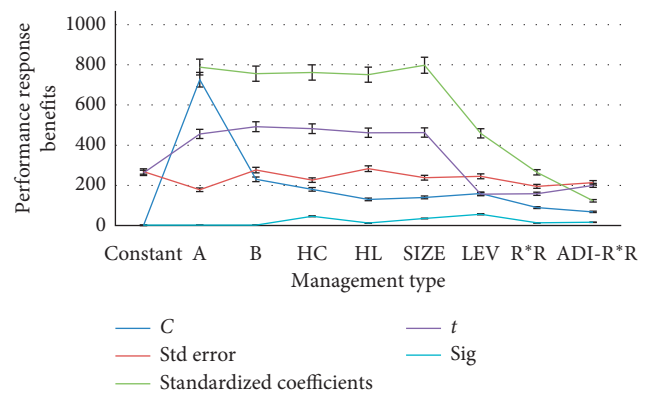


FIGURE 3: Regression analysis of human capital investment and enterprise performance.

TABLE 4: The more frequent flow of human capital.

| N  | Volatile factor | Pheromone concentration | Information quality | Valid data | Density |
|----|-----------------|-------------------------|---------------------|------------|---------|
| 10 | 3.82            | 1.47                    | 2.28                | 5.74       | 2.69    |
| 20 | 3.28            | 1.42                    | 2.47                | 1.19       | 4.07    |
| 30 | 5.17            | 3.75                    | 5.3                 | 3.31       | 5.07    |
| 40 | 6.2             | 1.72                    | 3.13                | 5.52       | 4.62    |
| 50 | 1.55            | 3.87                    | 2.74                | 2.79       | 1.84    |

TABLE 5: Employees' views on big data.

|            | Big data has a great impact on our lives (%) | Big data has a general impact on our lives (%) | Big data has a little impact on our lives (%) | Big data has a no impact on our lives (%) |
|------------|--|--|---|---|
| Proportion | 25   | 35   | 19  | 21  |

As shown in Table 5, these data and information are multidimensional. Companies need to process and analyze the information of these data to determine the current total number of company advantages in terms of employee composition, so as to distinguish between the company and competitors, especially the company's deficiencies in personnel structure. Prepare human resource planning; second, predict the company's human resource potential in the forecast phase.

The data in Figure 4 show that most people involved in human resource planning, including human resource managers and employees, do not generate much data awareness. Everyone says that they know they are in the era of big data, but they feel that big data is very far away from their life and work. 19% of the respondents think that big data has a general impact on their work and life, and 21% of them think that big data will not affect their work and life.

As shown in Table 6, the reality is that big data has penetrated into life and work, and human resource planning is no exception in all aspects. For the personnel involved in human resource planning, it is very important to cultivate a high degree of data awareness. Only when the relevant staff have a high degree of data awareness can they point out the value and influence of data in human design and make reasonable human resource planning according to the value reflected by big data. In this process, most people are managers and growers of enterprise employees.

As shown in Table 7, only when they take the lead in forming data awareness can they have a direct impact on other employees and promote the development of employee awareness in the whole enterprise. The formation of enterprise's overall consciousness can promote the digitalization process of enterprises and provide digital support for relevant personnel to predict the distribution of labor supply and demand.

**4.5. Market Use of Human Resource e-HR System.** With the development of human resource information software market, a certain scale has been formed. Domestic enterprises have a better understanding of the culture of domestic enterprises, and the user interface is more suitable for domestic enterprises. Especially after China's accession to the WTO, the software industry has sprung up like mushrooms. As shown in Figure 5, UFIDA, Kingdee, Qizheng, Neusoft, Jinyikang, and so forth have been recognized by domestic

enterprises, and they are ahead of foreign manufacturers in terms of cost performance.

As shown in Figure 6 and Table 8, although Microsoft, Oracle as the representative of the company, has been in the software industry for decades, its products are more mature and security is also higher; especially with Oracle database as the representative, open-source, security-incomparable, China's large enterprises generally choose Oracle as their own enterprise database. However, the price of foreign software industry is relatively high, and the maintenance cost is very high, and the interface does not necessarily conform to the norms and habits of Chinese enterprises. Therefore, the choice of information system should be combined with the enterprise itself rather than blindly choosing foreign software enterprises. According to the relevant statistical industry survey, most of China's enterprises have not yet introduced the information system for management. The figure below reflects the general trend of Chinese enterprises in choosing information system.

We can see clearly from Figure 7 and Table 9, where enterprises choose e-HR system, that the first consideration is the development stage of enterprise informatization. If the enterprise is just in the initial stage of informatization, it should choose a simple and easy-to-understand e-HR system for basic information operation. If enterprises do not require high degree of informatization, they only need simple employment data and choose e-HR system, which is easy to operate. If the enterprise develops into a medium-sized enterprise, is cross-regional, and has a large number of basic employees, it should choose e-HR system, which can process important data of the company and has additional functions.

As shown in Figure 8 and Table 10, if the enterprise is a large enterprise or requires a high degree of informatization due to the nature of the industry, it is necessary to select the e-HR system with the highest security level, which can process large data and provide data mining and data analysis. In a word, enterprises should choose e-HR system according to the development stage and scale of the enterprise and choose the e-HR system suitable for themselves.

**4.6. Human Capital Allocation in X Software Company.** As shown in Figure 9 and Table 11, X software company is a high-tech R&D company in the period of rapid

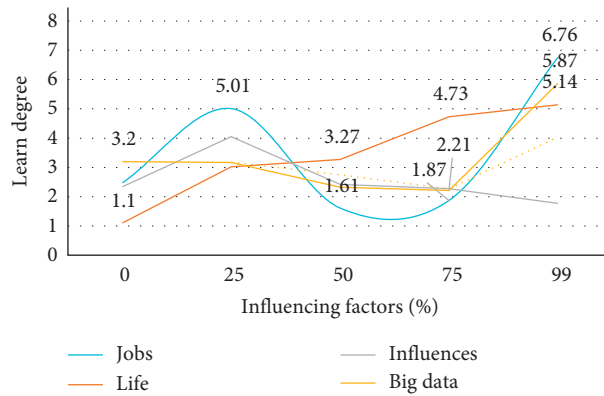


FIGURE 4: Employees' views on big data.

TABLE 6: The reality is that big data has penetrated into life and work.

| N  | Multicast routing | Convergence speed | Best ant colony | Ant colony member | Growing tree | Basic ant colony |
|----|-------------------|-------------------|-----------------|-------------------|--------------|------------------|
| 10 | 1.29              | 1.86              | 0.57            | 1.33              | 0.61         | 1.48             |
| 20 | 1.85              | 1.75              | 3.03            | 2.78              | 3.5          | 1.49             |
| 30 | 4.15              | 3                 | 3.42            | 3.31              | 3.58         | 5.82             |
| 40 | 1.47              | 2.28              | 5.74            | 2.69              | 2.53         | 5.35             |
| 50 | 5.62              | 5.28              | 9.16            | 6                 | 6.11         | 11.17            |

TABLE 7: Association of characteristic parts in complex systems.

| Num. | Group optimization | Ant colony optimization | Basic | Node | Effectiveness | The complexity |
|------|--------------------|-------------------------|-------|------|---------------|----------------|
| 10   | 2                  | 1.32                    | 1.45  | 0.86 | 0.08          | 0.57           |
| 20   | 3.19               | 3.16                    | 3.41  | 3.99 | 2.29          | 1.81           |
| 30   | 4.9                | 4.95                    | 3.71  | 2.55 | 2.71          | 5.13           |
| 40   | 4.54               | 2.07                    | 3.52  | 2.13 | 1.22          | 5.15           |
| 50   | 1.55               | 4.22                    | 2.53  | 1.42 | 4.1           | 1.06           |

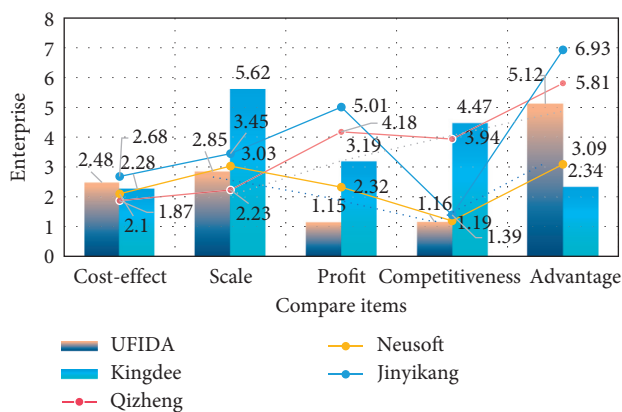


FIGURE 5: Competitive advantages of different companies.

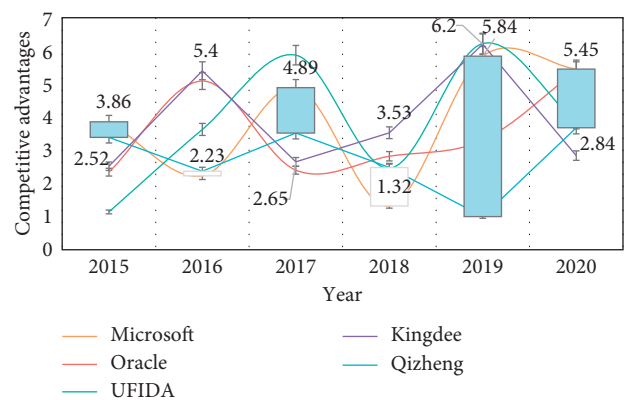


FIGURE 6: Competitive advantages of Microsoft and Oracle.

development. The staff is mainly composed of high-quality students, and the age structure is relatively young. The average age of the company's employees is about 25 years. The company has more than 90% of the total number of employees with certificates. According to the company's

technological innovation strategy, X software company has launched a large-scale recruitment plan.

As shown in Figure 10, the company's large amount of funds is used to attract knowledge-based and technical talents, introduce and improve the quality of employees, and

TABLE 8: Market utilization rate of e-HR system.

| Software type | UFIDA (%) | Kingdee (%) | Oracle (%) | Neusoft (%) | Sap (%) | Other (%) |
|---------------|-----------|-------------|------------|-------------|---------|-----------|
| Proportion    | 5         | 15          | 10         | 9           | 15      | 6         |

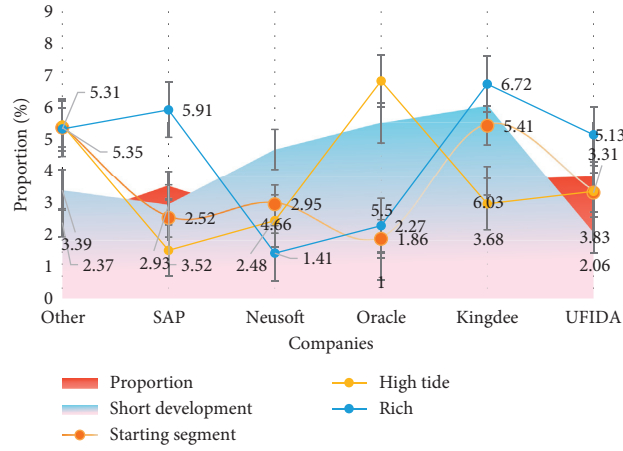


FIGURE 7: Market utilization rate of e-HR system.

TABLE 9: Market utilization rate of e-HR system data.

| Item    | Proportion | Starting segment | Short development | High tide | Rich |
|---------|------------|------------------|-------------------|-----------|------|
| Other   | 2.37       | 5.35             | 3.39              | 5.43      | 5.31 |
| SAP     | 3.52       | 2.52             | 2.93              | 1.5       | 5.91 |
| Neusoft | 2.48       | 2.95             | 4.66              | 2.43      | 1.41 |
| Oracle  | 1          | 1.86             | 5.5               | 6.82      | 2.27 |
| Kingdee | 3.68       | 5.41             | 6.03              | 2.96      | 6.72 |
| UFIDA   | 3.83       | 3.31             | 2.06              | 3.35      | 5.13 |

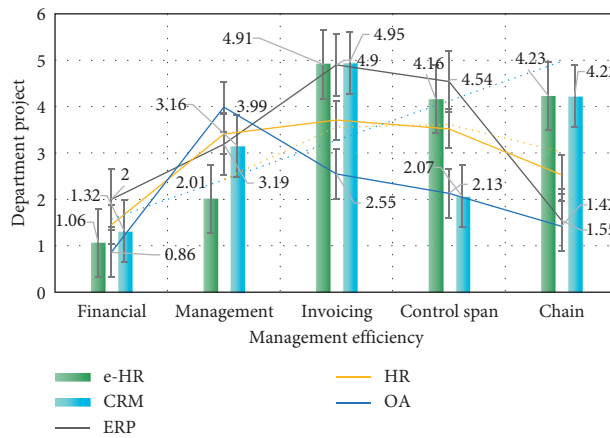


FIGURE 8: The relationship between company management tools and benefits.

TABLE 10: Relationship data between company management tools and benefits.

| Item         | e-HR | ERP  | CRM  | HR   | OA   |
|--------------|------|------|------|------|------|
| Financial    | 1.06 | 2    | 1.32 | 1.45 | 0.86 |
| Management   | 2.01 | 3.19 | 3.16 | 3.41 | 3.99 |
| Invoicing    | 4.91 | 4.9  | 4.95 | 3.71 | 2.55 |
| Control span | 4.16 | 4.54 | 2.07 | 3.52 | 2.13 |
| Chain        | 4.23 | 1.55 | 4.22 | 2.53 | 1.42 |

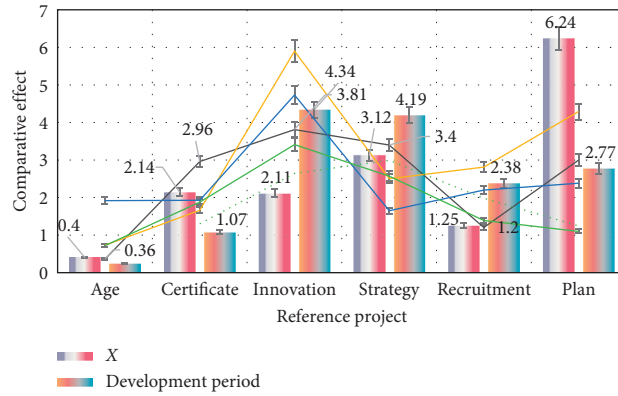


FIGURE 9: The relationship between company management and the quality of members.

TABLE 11: The influence of staff management and member quality on company benefit.

|             | X    | D    | Development period | Staff composition | High quality | Software |
|-------------|------|------|--------------------|-------------------|--------------|----------|
| Age         | 0.4  | 0.36 | 0.25               | 0.73              | 1.92         | 0.72     |
| Certificate | 2.14 | 2.96 | 1.07               | 1.66              | 1.93         | 1.88     |
| Innovation  | 2.11 | 3.81 | 4.34               | 5.9               | 4.73         | 3.41     |
| Strategy    | 3.12 | 3.4  | 4.19               | 2.5               | 1.65         | 2.57     |
| Recruitment | 1.25 | 1.2  | 2.38               | 2.81              | 2.2          | 1.39     |
| Plan        | 6.24 | 3    | 2.77               | 4.29              | 2.38         | 1.1      |

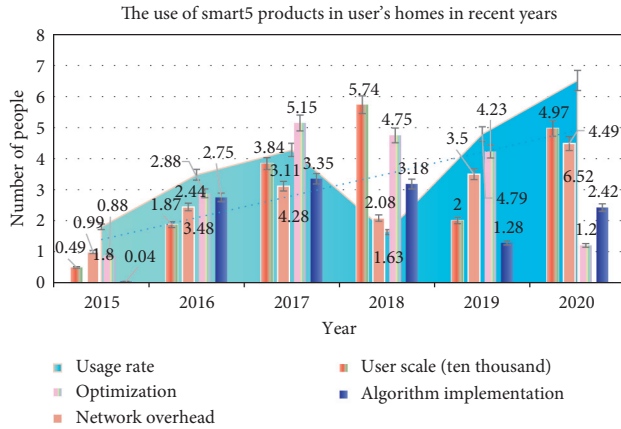


FIGURE 10: The stock of human capital of X software company.

enrich the company's structure. According to the human resource system disclosure, X software company has 2500 employees, of which more than 70% are R&D personnel. While the stock of human capital of X software is increasing rapidly, it should pay attention to maintaining a good structure of human capital. As a high-tech enterprise, the proportion of R&D personnel in the enterprise is about 70%, and the high proportion of R&D personnel ensures the development needs of Jinshan software products.

As shown in Figure 11, X software company attaches great importance to the training of its employees, and it takes one week every year for corporate culture training. The

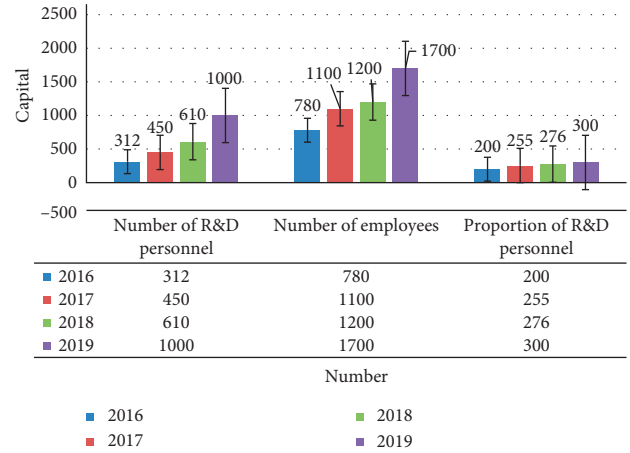


FIGURE 11: Human capital structure of X software company in 2016–2019.

senior management of the company introduces the development history and corporate culture of X software company to the employees.

4.7. Education Framework of Human Capital in High-Tech Enterprises. As shown in Figure 12, we can see that, among the high-tech enterprises, the proportion of high-tech enterprises with high-level education accounts for a large proportion, accounting for 45%, followed by master's degree accounting for 35% and undergraduate education accounting for 20%, which



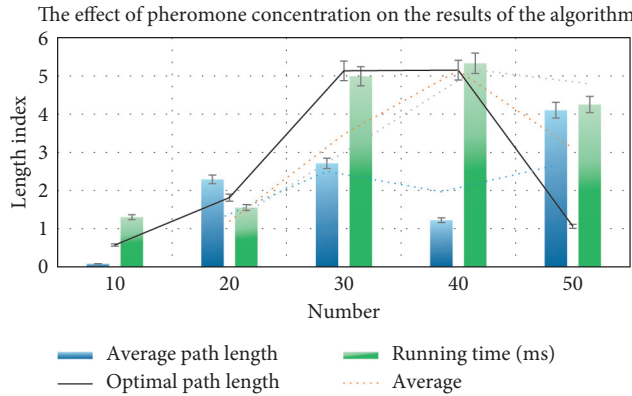


FIGURE 12: High-level education accounts for a large proportion.

TABLE 12: Education structure of human capital in high-tech enterprises.

| Education structure | Doctor's degree (%) | Master's degree (%) | Bachelor degree (%) | Other (%) |
|---------------------|---------------------|---------------------|---------------------|-----------|
| Tradition           | 15                  | 25                  | 30                  | 40        |
| Proportion          | 45                  | 35                  | 19                  | 1         |

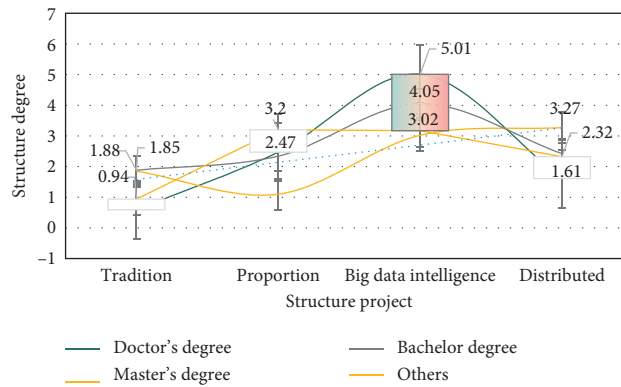


FIGURE 13: Education structure.

shows that the demand of high-tech enterprises for various talents is also increasing. In traditional enterprises, the proportion of high-tech enterprises with high degree is small, accounting for 15%, followed by master's degree accounting for 25% and undergraduate degree accounting for 30%.

As shown in Table 12 and Figure 13, a department takes over the technical training in the company and forms corresponding training programs, such as developing a small toy or small unit. Several people form a team. Many groups will develop at the same time. Finally, each group will produce products to show the comparison. For the training of senior officials, X software company will regularly select employees who have worked for more than two years to study for graduate students. The company will pay them more than half of their study expenses or send them to foreign countries to learn and exchange advanced technologies, so as to better contribute to the company's R&D.

4.8. Problems in Cost Control of Human Resource Use. As shown in Figure 14 and Table 13, from the overall level of the

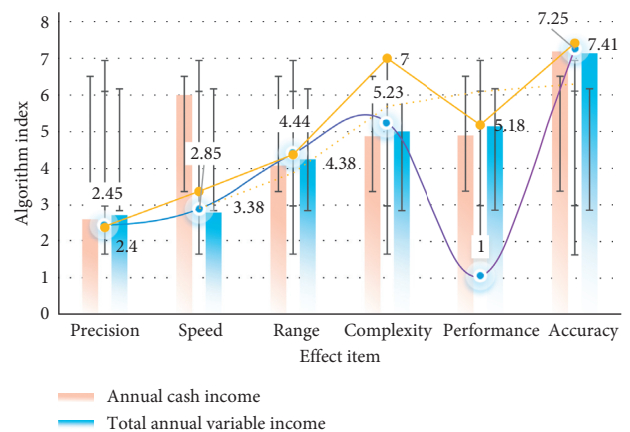


FIGURE 14: The proportion of human resource cost is increasing.

cost expenditure of high-tech enterprises in China, the proportion of human resource cost is increasing. The high-tech enterprises with strong economic strength can still afford it, but, for the small- and medium-sized high-tech

TABLE 13: Salary structure of human capital in high-tech enterprises.

| Occupation         | Annual cash income (%) | Total annual subsidy (%) | Total annual variable income (%) | Total annual benefits (%) | Total (%) |
|--------------------|------------------------|--------------------------|----------------------------------|---------------------------|-----------|
| The basic level    | 64                     | 7                        | 6                                | 23                        | 100       |
| Executive director | 62                     | 9                        | 9                                | 20                        | 100       |
| Management         | 65                     | 10                       | 13                               | 18                        | 100       |
| High level         | 68                     | 12                       | 5                                | 15                        | 100       |

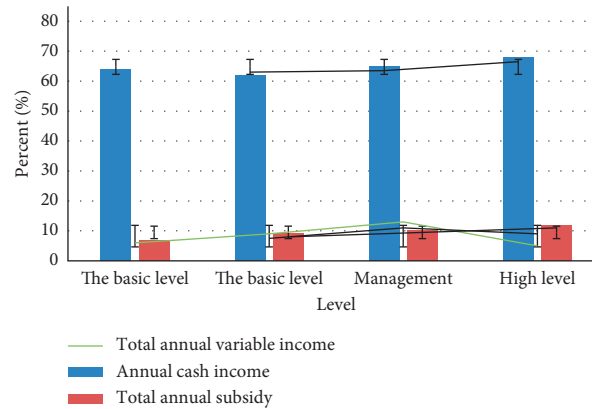


FIGURE 15: Salary structure of human capital in high-tech enterprises.

enterprises that have just started, this expenditure undoubtedly adds a huge economic burden to them.

From Figure 15, we can see that, in high-tech enterprises, whether at the grass roots, middle level, or senior level, the proportion of annual cash income, that is, employees' cash salary, is the highest. Such a large amount of capital investment will inevitably require the enterprise's distribution system to be transparent, open, and fair. However, the reality is that, compared with the market economy, there are still many drawbacks in the current salary system of some high-tech enterprises, and the salary distribution system is directly related to the expenditure of the use cost of human resources, so the salary system also directly affects the control of the use cost of human resources.

## 5. Conclusions

In the economic level, science and technology tell development today, network and intelligent are no longer a strange topic, and this era is a modern era. In the field of human resources, it is no longer an era of tarnishing exploration. Relying on big data technology, the research on human resources of high-tech enterprises has entered a climax. With the in-depth application of cloud computing and other technologies, the construction of high-tech enterprise informatization is accelerated. It is a long-term process to build and implement the big data analysis platform for human resources of high-tech enterprises based on cloud computing. With the development of human resource system to cloud information, more and more enterprises will apply the human resource big data analysis platform based

on cloud computing. Therefore, the human resource big data analysis platform can contain information of many enterprises. Different enterprises can compare their own information with the information publicly disclosed by competitive enterprises in real time and make favorable decisions when they know themselves and their competitors. At the same time, the government can also use the platform to remote audit different enterprises to facilitate government supervision.

Of course, the support based on big data technology is far from enough. Every company should pay attention to the training of employees and spend a lot of time and energy on the allocation of employees among different departments, including job needs, employee abilities, employee wishes, and salary distribution. There are too many factors involved, so allocating employees in human resources will encounter various problems. Moreover, due to the lack of timely and in-place communication, there is a phenomenon of "re-work" in the work. That is, after working for a period of time, the employees reflect that they are not suitable and need to be readjusted. Therefore, this has brought about great trouble to the human resources department and delayed the work progress of the company. The use of big data can greatly avoid such a thing, because, before the allocation of manpower, big data will have human resource assessment, which can adjust the special situation in advance until it is suitable. Therefore, big data can help human resource units to make good use of human resources.

Human resource information management is the trend of future development; it can promote the value-added of human resources, and human resources management has

been raised to a strategic position; the implementation of human resources information management can enhance the core competitiveness of enterprises, which has important strategic significance for the development of enterprises. For high-tech enterprises, the success of informatization depends on staff training to a large extent. In the process of training, training should be carried out at different levels. Senior leaders focus on decision-making ability and strategic management ability, middle-level functional managers focus on communication ability, and grass-roots employees focus on technical ability. To do a good job in training evaluation, information management based on big data technology is a long-term implementation process. Enterprises should continuously improve the system, and information management is also a process of full participation.

### Data Availability

No data were used to support this study.

### Conflicts of Interest

The author declares that there are no conflicts of interest.

### Acknowledgments

This work was supported by the Ministry of Education, Humanities and Social Science projects (18YJAZH128.)

### References

- [1] A. L. Ranft and M. D. Lord, "Acquiring new: the role of retaining human capital in acquisitions of high - tech firms," *The Journal of High Technology Management Research*, vol. 11, no. 2, pp. 295–319, 2020.
- [2] G. Bihua, "Cost accounting and management of R&D projects in high-tech enterprises," *Accounting Study*, vol. 2, no. 19, pp. 189–190, 2017.
- [3] S. Qin, "On human capital investment," *Economics and Management*, vol. 3, pp. 15–16, 2020.
- [4] H. Chen, B. Xuan, P. Yang, and H. Chen, "A new overhead crane emergency braking method with theoretical analysis and experimental verification," *Nonlinear Dynamics*, vol. 98, no. 3, pp. 2211–2225, 2019.
- [5] D. Sledgianowski, M. Gomaa, and C. Tan, "Toward integration of big data, technology and information systems competencies into the accounting curriculum," *Journal of Accounting Education*, vol. 38, no. 38, pp. 81–93, 2017.
- [6] F. You, C. Huang, and S. He, "Extended theoretical analysis method on the performance of high-efficiency power amplifiers by solving nonlinear waveform determination process," *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 27, no. 4, pp. 1–13, 2017.
- [7] Z. Wang, J. Yang, H. He et al., "The mechanism of politically-connected managers on technological innovation for small and medium high-tech enterprises - empirical evidence from listed firms in Chinese SMEs stock market," *International Journal of Service & Computing Oriented Manufacturing*, vol. 1, no. 2, pp. 178–195, 2017.
- [8] Z. L. Chen and Y. Q. Xiong, "Theoretical analysis for fabrication of two-dimensional photonic crystals by dual-beam holographic lithography method," *Rengong Jingti Xuebao/ Journal of Synthetic Crystals*, vol. 47, no. 1, pp. 97–101, 2018.
- [9] R. Wan, L. Guo, T. Li et al., "Theoretical analysis of lattice-mediated plasmon resonance using finite-difference time-domain method," *Journal of Nanoscience and Nanotechnology*, vol. 19, no. 1, pp. 40–46, 2019.
- [10] T. Philippon and A. Reshef, "Wages and human capital in the U.S. Finance industry: 1909–2006," *Quarterly Journal of Economics*, vol. 127, no. 4, pp. 1551–1609, 2017.
- [11] B. M. Fraumeni, M. S. Christian, and J. D. Samuels, "The accumulation of human and nonhuman capital, revisited," *Review of Income & Wealth*, vol. 63, no. 2, pp. 227–286, 2017.
- [12] J. Crespo Cuaresma, "Income projections for climate change research: a framework based on human capital dynamics," *Global Environmental Change*, vol. 42, pp. 226–236, 2017.
- [13] M. Do, L. Bungert, D. Cichon et al., "Localization of passive 3-D coils as an inverse problem: theoretical analysis and a numerical method," *IEEE Transactions on Magnetics*, vol. 56, no. 4, pp. 1–10, 2020.
- [14] G. Yang, G. Ma, and G. Zeng, "Theoretical analysis of generating two-stage non-diffracting beam based on axicon method," *Journal of Applied Optics*, vol. 39, no. 5, pp. 722–728, 2018.
- [15] J. Roufagalas, A. G. Orlov, and M. Bahmanioskooee, "Endogenous growth, human capital and the dynamic costs of recessions," *Journal of Economic Studies*, vol. 47, no. 2, pp. 264–285, 2020.
- [16] S. S. Omar, Z. Z. Ariffin, N. Mahadi, Z. Ahmad, and R. R. Sarah, "Employees' human capital: multiple-case studies of small manufacturing firms," *Advanced Science Letters*, vol. 24, no. 6, pp. 4644–4648, 2018.
- [17] P. Kumar and Gulshan, "Theoretical analysis of extreme wave oscillation in Paradip Port using a 3-D boundary element method," *Ocean Engineering*, vol. 164, no. 15, pp. 13–22, 2018.
- [18] R. A. Brymer, C. Chadwick, A. D. Hill, and J. C. Molloy, "Pipelines and their portfolios: a more holistic view of human capital heterogeneity via firm-wide employee sourcing," *Academy of Management Perspectives*, vol. 33, no. 2, pp. 207–233, 2019.
- [19] W. Gaertner and L. Schwettmann, "Burden sharing in deficit countries: a questionnaire-experimental investigation," *Series*, vol. 8, no. 2, pp. 1–32, 2017.
- [20] C.-Y. Liao, Y.-C. Wu, C.-Y. Chang, and C.-C. Ma, "Theoretical analysis based on fundamental functions of thin plate and experimental measurement for vibration characteristics of a plate coupled with liquid," *Journal of Sound and Vibration*, vol. 394, no. 4, pp. 545–574, 2017.
- [21] H. Xie, W.-I. Li, Z.-P. Yin, and H. Ding, "Variance-minimization iterative matching method for free-form surfaces-Part II: experiment and analysis," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 3, pp. 1192–1204, 2019.
- [22] L. Zattra and N. Donin, "A questionnaire-based investigation of the skills and roles of computer music designers," *Musicae Entiae*, vol. 20, no. 3, pp. 436–456, 2017.
- [23] T. Aburai, A. Okubo, D. Suzuki, and K. Takeyasu, "Bayesian network analysis for the questionnaire investigation on tourists' behavior," *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, vol. 31, no. 4, pp. 764–770, 2019.
- [24] B. Zou, J. Dong, Y. Yao, and Y. Jiang, "A detailed study on the optical performance of parabolic trough solar collectors with

Monte Carlo Ray Tracing method based on theoretical analysis,” *Solar Energy*, vol. 147, no. 4, pp. 189–201, 2017.

- [25] T. Kenzaka, A. Kumabe, K. Kosami et al., “Bacteriological testing and recurrence prevention efforts in the diagnosis and treatment of nursing- and healthcare-associated pneumonia and aspiration pneumonia: a questionnaire survey of hospitals across Japan,” *Respiratory Investigation*, vol. 56, no. 2, pp. 150–157, 2018.

## Research Article

# Development and Supervision of Robo-Advisors under Digital Financial Inclusion in Complex Systems

Wensheng Dai 

Financial School, China Financial Policy Research Center, International Monetary Institute, Renmin University of China, Beijing 100872, China

Correspondence should be addressed to Wensheng Dai; daiws@ruc.edu.cn

Received 18 October 2020; Revised 17 December 2020; Accepted 4 January 2021; Published 28 January 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Wensheng Dai. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of the market economy, there are more and more projects in the financial industry, and their complexity and technical requirements are getting higher and higher. The development of computer technology has promoted the birth of robot consultants, and it is of great significance to use robot consultants to manage and supervise financial industry projects. In order to further analyze the development and supervision of robo-advisors under the digital inclusive financial system, this paper uses complex systems and clustering algorithms as technical support to carry out research. First, the traditional K-means algorithm is used to select the initial clustering center, to improve the noise and outlier processing capabilities, and to build a data mining system based on the improved algorithm. Then, a product design model for robo-advisors is built and the risks of robo-advisors are analyzed from three aspects: technology, market, and law. Analyzing the performance of the improved K-means algorithm, in the operation of the experimental dataset B, the accuracy of the clustering result after 6 iterations reached 97.08%, which shows that the algorithm has good performance. During the trial operation of the data mining system, the four types of customers of financial institutions were accurately clustered, and it was concluded that the main type of customers who brought benefits to financial institutions was high-income customers accounting for 10.75%. Robo-advisory product models are used to build five risk-level investment portfolios and conduct risk backtests. Except for the growth and income portfolio, other portfolios have consistently outperformed the performance benchmark during the analyzed time period. Running the research system of this paper in a financial institution, comparing the capital budget before and after the operation, found that the system can improve the accuracy of the budget and reduce the risk of the robo-advisor for the financial institution.

## 1. Introduction

**1.1. Background Significance.** In the operation of financial activities, there are many projects and tasks in parallel, which brings great impact to the traditional financial management concept. As an important application of financial technology in the field of wealth management, the mode of intelligent investment adviser is more complex. In the era of big data, there are many problems in the development and supervision of intelligent investment advisers [1]. Complex systems can be said to be all over every corner of daily life. Complexity science is an emerging research form that reveals the operation laws of complex systems [2]. The development and supervision of robo-advisors in the digital age is also an extremely complex research object. Therefore,

it is a unique and meaningful new idea to study the development and supervision of digital inclusive finance and robo-advisors from the perspective of complex systems.

**1.2. Related Work.** Complex systems have become the focus of research in various fields due to their complexity and extensiveness. Lehuta et al. focused on handling uncertainties by optimizing model complexity for management goals and technical issues to increase confidence in complex system models. They reviewed how the complex system model fits into the existing institutional and legal environment of the current European fishery decision-making framework [3]. Although their research is of reference significance, their research methods lack innovation.

Inclusive finance plays an important role in improving the income gap and improving the living standards of the poor and disadvantaged groups, so it is the object of key research. Yan et al. studied the impact of digital financial inclusion (DFI) on the stabilization of household consumption in China. They used the data from the two “Chinese Family Forum” studies from 2010 to 2016. They divided household income shocks into permanent and temporary parts and assessed whether digital financial inclusion can help families resist income shocks [4]. Their research data are very representative but lack certain accuracy in processing the data. The risk analysis of robo-advisors has always been the focus of attention in the financial field. Jung et al. determined the needs of robo-advisors, derived design principles, and evaluated it through algorithm iterations in a controlled laboratory study [5]. Their research has given us a deeper understanding of robo-advisors, but they have not made constructive suggestions for the improvement of its supervisory system.

*1.3. Innovative Points in This Paper.* In order to build a more complete robo-advisory supervision system, reduce risks, and improve the digital level of inclusive finance, this paper studies the development and supervision of robo-advisors based on complex systems and clustering algorithms. The innovations of this research are as follows. (1) Improve the traditional K-means algorithm, optimize the selection of its initial clustering center, reduce the influence of noise, and improve the processing ability of isolated points. (2) A data mining system is constructed based on the improved algorithm. The functions of the system include opening files, importing data, data preprocessing, data clustering, and result query. (3) This paper constructs the product design model of intelligent investment consultant and uses the model to construct five risk-level portfolios for risk back-testing. This paper analyzes the risks of intelligent investment advisers from three aspects of technology, market, and law and puts forward suggestions to improve the supervision of intelligent investment advisers.

## 2. Complex Systems and Technologies Related to Digital Inclusive Finance

### 2.1. Complex System

*2.1.1. Characteristics of Complex Systems.* Complex systems exist in every corner of human life. Ecosystem, population system, and global economic system belong to the category of the complex system. They all have the same characteristics as the complex system. Complex systems are systematic first, which is not the superposition of simple systems and organizations. Therefore, it is not possible to study complex systems with traditional system analysis methods [6, 7]. The elements of a complex system are in a nonlinear relationship. Simple partial stacking cannot represent the whole. The local laws are not the same as the overall laws. Therefore, a new system theory is needed to consider the logical relationship between complex systems.

Complex systems are also hierarchical and interactive. The hierarchical nature of the complex system is mainly embodied in the nested relationship of different levels of interconnectedness [8]. Therefore, in the research of complex systems, it is necessary to update the traditional concept of hierarchy and analyze the research objects from the level of complex system theory. Complex systems and the external environment always interact. Different complex systems together form a larger and more complex system. When studying a complex system, we must fully consider the internal environment and external environment, study the information exchange between them, and consider the self-adjustment of the complex system in the complex external environment.

Complex systems have emergence and development. Complex systems are composed of various subsystems and local subsystems which are composed of various combinations and correlations. If the format and functional structure of the subsystem and the local subsystem are different, the complex system will no longer be the sum of the subsystem functions [9]. Complex systems may have a variety of new features, so when studying complex systems, we must consider the original features and the various new features that may appear. The self-renewability of complex systems is mainly reflected in the continuous development of the system, which is also the fundamental reason for biological evolution and the development of human society. Complicated systems become intelligent due to internal hierarchical, systematic and external interactivity, and emergence, so they can adapt to the needs and changes of the environment.

*2.1.2. Agent Complex System.* The complex adaptive system is based on the characteristics of the complex system and further develops the Agent theory. Agent is an independent individual or subsystem in a complex system, with a life cycle, which can perceive and adapt to the environment, run autonomously in the environment, and even change the environment. The structure of Agent generally includes environment perception, reasoning, control decision-making, knowledge base, and communication [10].

Agent has the characteristics of autonomy, social ability, initiative, learning, and adaptability [11]. Agents can use their own state and knowledge to make decisions independently, without relying on outside help. Agents can achieve a certain degree of communication, negotiate and cooperate to resolve conflicts, and complete complex tasks. Agents can judge their own situation according to the external environment and actively make choices that are beneficial to themselves at the right time. Agents can also continue to learn, adjust their own state and behavior, and adapt to the constantly changing external environment.

A single Agent has autonomous capabilities to a certain extent, but a single Agent cannot complete work in a complex and changeable environment. A multiagent system emerged at the historic moment, in which each Agent can communicate with each other. There are two or more Agents in a multiagent system, each with autonomy but limited

capabilities [12]. The multiagent system does not have complete global control. A single Agent has its own judgment and status. The calculation of the entire system is asynchronous, concurrent, or parallel. The coordination methods between Agents in the multiagent system are classified as follows.

As shown in Figure 1, the coordination methods of multiagent systems can be divided into two categories: explicit coordination and implicit coordination. The explicit coordination includes complete centralized coordination, complete distributed coordination, and centralized and distributed combined coordination, and implicit coordination includes social rules and filtering strategies.

*2.1.3. Evolutionary Dynamics of Complex Systems.* There are some common dynamic behaviors in the evolution of complex systems. Under the action of dynamics, complex systems can evolve and develop continuously. It is also feasible to use the dynamic behavior of individuals to illustrate the evolution of the entire complex system. Individual-based dynamic processes are usually defined by strategy update rules. The strategy update rule is based on factors such as environment and information to update and select the benchmark of the game strategy adopted by the individual in the system.

The evolutionary power strategy update rules of complex systems can be divided into two categories: imitation-learning strategy update rules. In this process, only the original strategy is imitated without innovation; the strategy update rules of the optimal response process include the optimal response process generated by the new strategy. In the process of evolution, individuals can use different types of strategies to update rules or use the same strategy to update rules [13]. There are two types of update methods: synchronous update and asynchronous update. Synchronous update refers to the synchronous update of the entire overall population in each individual time step. Asynchronous updates are updated in batches in a specific order.

In a complex system, individuals use a series of different dynamic behaviors in the development process according to different strategic update rules so that the evolution of the complex system shows the common characteristics of a series of complex systems. The dynamic behavior and dynamic factors of specific complex system evolution are also involved. Most of them start from the perspective of game theory and analyze the evolution of complex systems represented by individual competition, cooperation, and learning interactions in complex systems.

## 2.2. Digital Inclusive Finance

*2.2.1. Content Framework of the Inclusive Financial System.* The connotation of inclusive finance reflects fairness, comprehensiveness, and profitability, so inclusive finance has the characteristics of inclusive and sustainable development [14]. The complete framework of the inclusive financial system considers that large-scale groups excluded by traditional financial institutions want to obtain financial

services and benefit from them. Financial services must be integrated into the customer, micro, meso, and macro levels of the financial system. The framework of the inclusive financial system is as follows.

As shown in Figure 2, the customer level refers to the key service targets of inclusive finance, including vulnerable groups excluded by formal finance. The micro level refers to providers of financial services, including formal and informal financial institutions, as well as other types of financial institutions in the middle. The meso level refers to financial infrastructure and related services, including basic financial facilities, a series of auxiliary services that expand the scope of services, reduce costs, and promote transparency [15]. The macro level refers to the appropriate development environment, including the macroeconomic environment, institutional environment, and supervision and management environment.

The supply of the inclusive financial system is composed of commercial, policy, and cooperative financial institutions at the mature stage, and its structure is as follows.

As shown in Figure 3, commercial, policy, and cooperative financial institutions together form the ternary supply structure of the inclusive financial system at the mature stage [16]. They complement each other. Policy financial institutions provide policy guidance, and commercial financial institutions provide capital flows. Cooperative financial institutions provide necessary subsidies.

### 2.2.2. Development Stage of Digital Inclusive Finance.

Digital inclusive finance extends and expands the meaning of financial services by using digital technologies such as the Internet and big data, expands the coverage of its services, and reduces the threshold and cost of financial services [17]. Moreover, digital financial inclusion can promote information sharing and reduce the degree of information asymmetry. The development stages of digital financial inclusion are as follows.

As shown in Figure 4, the development of digital financial inclusion has gone through four stages. The initial form of microfinance was very simple, and then new business models such as the individual farmer household loan system and the installment payment system appeared. Although it overcomes the moral crisis brought about by information asymmetry to a certain extent, the scope is limited. And, its funding sources are mainly foreign aid funds provided by the government, international aid, and nonprofit organizations. Once the foreign aid funds are reduced, it will not be possible to continue lending, and sustainable development will face challenges [18]. The development of microfinance is more sustainable and diverse. Sustainability is reflected in the expansion of its business into deposits, wealth management, and other fields, without relying on the financial support of the government, other organizations, and individuals and financial sustainability. Diversity is reflected in the diversification of institutions, businesses, and service targets. Inclusive finance not only provides financial products and services to the poor and

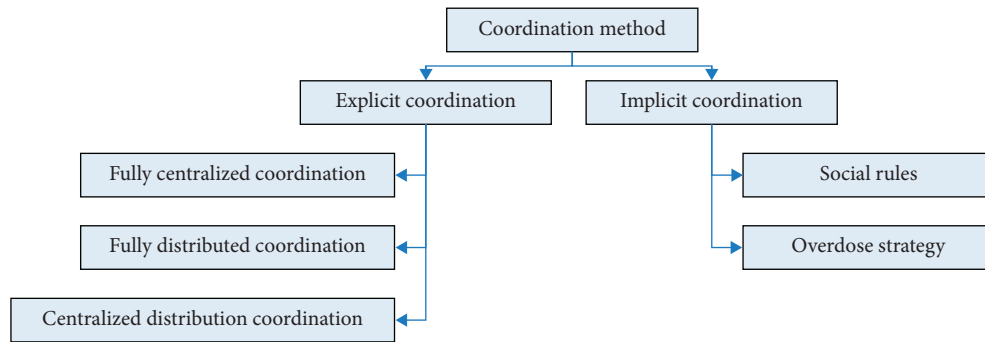


FIGURE 1: Coordination method of the multiagent system.

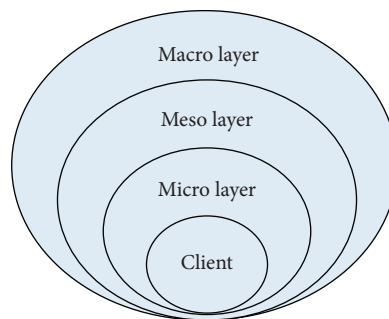


FIGURE 2: Framework structure of the inclusive financial system.

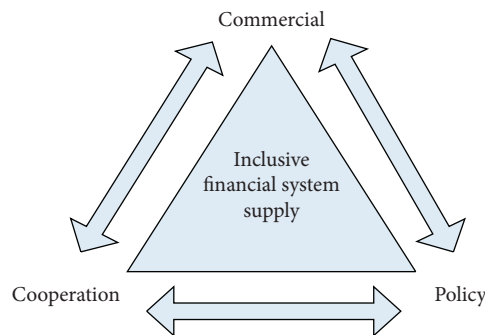


FIGURE 3: Supply structure of the inclusive financial system.

vulnerable groups but also promises to meet the diverse financial needs of all classes and groups in the society and encourages the wide participation of various financial institutions in the society. Digital finance is a new financial model that not only reduces transaction costs and information asymmetry but also expands the possibility of a series of transactions.

*2.2.3. Threshold Mechanism for the Development of Inclusive Finance.* The development of inclusive finance needs to face its threshold mechanism. In order to ensure profitability, financial institutions will set thresholds for mortgages and other services. Only when customers pay for financial services can they enjoy the services. In the mature stage of

financial development, low-income groups who were originally excluded from the financial service system can gradually accumulate wealth to enjoy financial services, which will narrow the investment income gap between different groups [19]. It is possible to establish a theoretical analysis framework for the impact of financial development on the income gap from the perspective of the production cycle of the two families.

Assuming that there are rich and poor families, the production cycle of the two families is the same, and the financial service lending thresholds are the same; we can deduce the wealth accumulation process of the two families in the context of the borrowing threshold constraints. A poor family can only save value first to maintain simple reproduction. The initial wealth of the family is as follows:



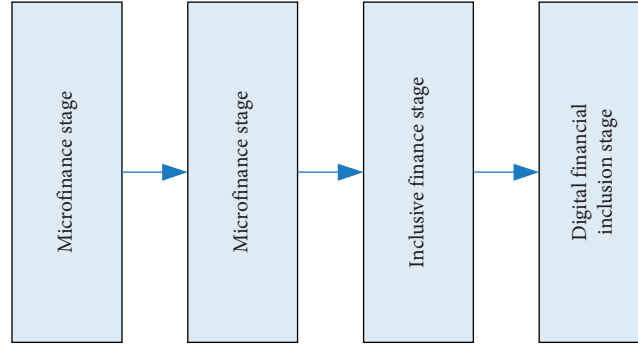


FIGURE 4: Development stage of digital financial inclusion.

$$C = K - D + (\alpha + e)Y_0. \quad (1)$$

Among them,  $K$  refers to the family's external endowment in the  $t$  period,  $\alpha$  refers to the proportion of the family's current wealth left to the next generation in each period of wealth distribution, and the remaining part of the current wealth for consumption is  $D$ .  $Y_0$  is the initial wealth from the family inheritance of the previous generation. After the preservation and saving with an interest rate of  $e$ , the wealth of the family in period  $t$  is as follows:

$$C_t = K + W_{t-1} + eC_{t-1} - D. \quad (2)$$

Among them,  $W_{t-1}$  refers to the inheritance of the previous generation family, so the inheritance left to the next generation by the family is as follows:

$$W_t = \alpha(K + W_{t-1} + eC_{t-1} - D). \quad (3)$$

From formulae (1)–(3), the convergence value of the wealth of the poor family can be derived:

$$C_\infty = \frac{(G - D)}{(1 - \alpha - e)}. \quad (4)$$

It can be seen from formula (4) that because it is difficult for poor families to cross the borrowing threshold, they are easy to fall into the “poverty trap” without the intervention of external forces. For wealthy families, their initial wealth is relatively abundant, and they can cross the borrowing threshold and obtain credit support. Therefore, the wealth of the family in period  $t$  is as follows:

$$C_t = K + W_{t-1} + S - M \times (1 + i) - H - D. \quad (5)$$

Among them,  $M$  is the amount of credit support,  $S$  is the income of each period after the credit support is invested in the project, and  $H$  is the cost of borrowers entering the financial market. In addition, the return on investment projects must be greater than the cost of the loan.

$$B = K + S - M \times (1 + i) - H - D. \quad (6)$$

According to formula (6), we can calculate the legacy of wealthy families to the next generation:

$$W_t = \alpha C_t = \alpha(W_{t-1} + B). \quad (7)$$

The convergence value of the wealth of the rich family can be deduced according to formulae (5)–(7):

$$C_\infty = \frac{B\alpha}{(1 - \alpha)}. \quad (8)$$

From formula (8), it can be seen that increasing the proportion of legacy left to the next generation and reducing the cost of entering the financial market can increase the wealth of wealthy families.

### 2.3. Related Technologies of Financial Data Mining

**2.3.1. Data Mining Technology and Data Storage Technology.** Data are the basis of data mining, and various algorithms are the means of data mining. The ultimate goal of data mining is to obtain the knowledge contained in the data to assist analysis and decision-making [20]. The mining of financial data is to promote the development of financial economy. The content of data mining includes things and dimensions, distribution and relationships, description and prediction, and phenomena and knowledge [21]. The main techniques of data mining include forecasting, clustering, association rules, and time series analysis.

MySQL is a relational database management system. The relational database uses the structured query language SQL as the most commonly used standard language to store data in various tables and access [22]. It can be used as individual application in the network environment of client and server. It can also be incorporated into other software as a library to provide multi-language support. MySQL database has good versatility and can improve storage efficiency.

Memcached is a high-performance distributed memory object caching system, which can reduce the load of the database, reduce the frequency of database access, and increase the high load of dynamic and data-driven websites [23]. Memcached has a simple protocol and built-in memory storage method. The data management access between the data layer and the service layer can cope with the emergence of high concurrent access and high concurrent query scenarios.

2.3.2. *Bayes Classifier.* The Bayes principle describes the probability of event  $Y$  when event  $X$  occurs. The Bayes formula is as follows:

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}. \quad (9)$$

In order to adapt formula (9) to the entire sample space, it is necessary to derive the total probability formula on the basis of conditional probability. Assuming that the sample space is  $S$  and the remaining part of the sample space except event  $X$  is  $X'$ , the total probability formula is the conditional probability formula summarized in this sample space.

$$P(Y) = P(Y \cap X) + P(Y \cap X'), \quad (10)$$

$$P(X \cap Y) = P(Y|X)P(X), \quad (11)$$

$$P(Y) = P(Y|X)P(X) + P(Y|X')P(X'). \quad (12)$$

We can derive formula (12) from formulae (10) and (11). This is the total probability formula.  $P(X)$  is the prior probability,  $P(X|Y)$  is the posterior probability, and  $P(X|Y)/P(Y)$  is the likelihood function.

2.3.3. *K-Means Clustering Algorithm.* The K-means clustering algorithm is very common in the field of data mining. In this algorithm, K clustering centers are used to describe the clustering results. When the clustering objective function is known, the clustering analysis can be carried out by iterative updating to get the clustering results [24]. The schematic diagram of the K-means clustering algorithm process is shown in Figure 5:

The K-means clustering algorithm expresses the similarity measure between different clustered objects and data through distance. The performance evaluation method of the clustering result of this algorithm is usually the square of the mean square error [25]. After the iterative calculation, the objective function value of the algorithm will decrease, and the operation of the algorithm is a continuous iterative process.

The K-means clustering algorithm can process image and text features and has high stability and scalability. The clustering effect is good, and the result is more intuitive and easy to understand. When using an approximate algorithm to process datasets, as long as the target datasets are independent of each other, there is no need to restrict the scope of the dataset. However, the K-means clustering algorithm has limitations in practical applications, and the K value will directly affect the clustering effect. It is easy to be disturbed by local noise in the iterative process, which makes the result bias.

### 3. Experiments on Algorithm Optimization and Robo-Advisor Development and Supervision

#### 3.1. Improvement of K-Means Clustering Algorithm and Construction of the Data Mining System

3.1.1. *Improve the Selection of Initial Cluster Centers.* The key technologies to improve the K-means clustering algorithm include clustering information resource sharing technology based on the interoperability meta-model framework and cluster automatic detection technology based on multiengine fusion. The square error and the benchmark function can be used to achieve cluster analysis of datasets with large differences in shape. The most convenient method is to randomly select multiple different initial values to execute the algorithm to reduce the dependence of the K-means clustering algorithm. According to the processing method of the representative points, the ideal clustering result is obtained by selecting the global best solution in the clustering result. When choosing the initial clustering center of the clustering algorithm, it is necessary to improve the selection of the initial clustering center.

In order to improve the stability of K-means clustering algorithm results, it is necessary to select data samples according to the characteristics of data distribution. In the process of determining the initial cluster center, it can reflect the characteristics of the initial data, and the data will not be distorted after sampling.

3.1.2. *Improve Noise and Outlier Processing Capabilities.* Perform clustering according to the steps of the traditional K-means clustering algorithm, delete outliers, and obtain a new dataset. Select the data object corresponding to the smallest density value from the set  $S$  as the initial cluster center. When the distance between each point and the data and the remaining points is greater than the sum of the distances between the data object and all points, this point is treated as an isolated point.

In order to reduce the interference influence of the clustering effect of the K-means clustering algorithm, it is necessary to calculate the sum of the distance between each point and the data and the remaining points and the distance between the data object and all points.

$$Q_i = \sum_{j=1}^e \sqrt{\sum_{d=1}^e (x_{id} - x_{jd})^2}, \quad (13)$$

$$H = \sum_{i=1}^e \frac{Q_i}{e}. \quad (14)$$

Among them,  $e$  and  $d$  are the sample data and data dimensions, respectively;  $i$  is all points in the set; and  $j$  is a constant.

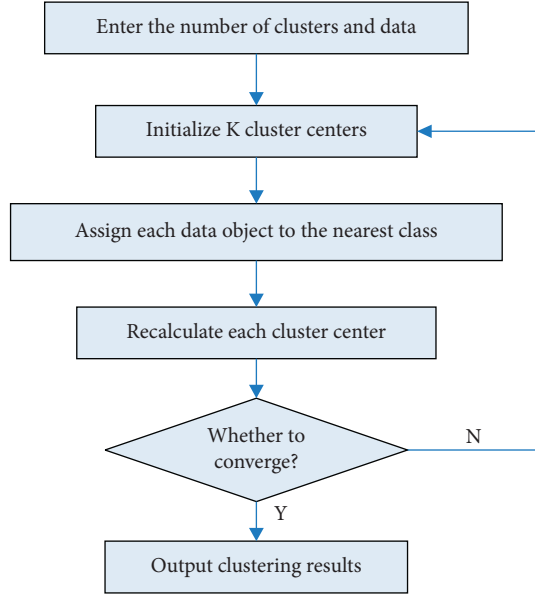


FIGURE 5: Schematic diagram of K-means clustering algorithm process.

**3.1.3. Data Mining System Construction.** There are five functions of the data mining system, which are file opening, data importing, data preprocessing, data clustering, and result query. The overall program flow of the system is as follows.

As shown in Figure 6, the user can specify the stored data file according to the flow instructions and then preprocess the data. After the data are stored in the database, users can choose whether to generate data or not and can query the data.

**3.2. Robo-Advisor Product Design Model.** Assuming that the asset portfolio  $R$  selected by the investor contains  $M$  different asset targets, the capital weight for investing in the  $n$ -th asset is  $X_n$ , and the expected return calculation method of portfolio  $R$  is shown in the following formula:

$$Q_R = \sum_{n=1}^M X_n Q_n. \quad (15)$$

The variance calculation of portfolio  $R$  returns is shown in the following formula:

$$\sigma_R^2 = \sum_{n=1}^M X_n^2 \sigma_n^2 + \sum_{n=1}^M \sum_{\substack{j=1 \\ j \neq n}}^M X_n X_j \sigma_{nj}. \quad (16)$$

For each asset target,  $\sigma_n^2$  is a diversifiable risk, but  $\sigma_{nj}$  is a nondiversifiable overall market risk. Then, when the target expected rate of return of the investment portfolio is determined, the investment portfolio with the least risk has the largest rate of return. The expression is shown in the following formula:

$$\begin{aligned} & \min \alpha^T \sum \alpha \\ \text{u.c.} & \begin{cases} \alpha E(Q_R) = \eta \\ 0 \leq \alpha_n \leq 1 \\ 1^T \alpha = 1 \end{cases} \end{aligned} \quad (17)$$

Among them,  $\eta$  is the expected rate of return of the investment portfolio target and  $\alpha$  is the combination weight of the asset and the covariance matrix between the asset targets.

**3.3. Risk Analysis of Robo-Advisor.** The risk analysis of robo-advisors can be conducted from three aspects: technology, market, and law. Technical risks can be divided into internal and external. The external is cyber risk, and the internal is algorithmic defects and operational risks. The investment effect of market risk on robo-advisors needs to be tested by the market. The homogeneity of robo-advisory algorithms may have an intensifying effect on market resonance, thereby exacerbating the procyclicality of investment behavior. The complexity of robo-advisor business is directly proportional to the correlation between institutions. Legal risks include the contradiction between operational intelligence and fiduciary obligations and the contradiction between the professionalism of algorithms and regulatory methods.

## 4. Discussion on Robo-Advisor Development and Regulatory Issues

### 4.1. Algorithm Performance and System Operation Effect

**4.1.1. Algorithm Performance.** In order to verify the effectiveness of the improved K-means clustering algorithm, three datasets (A, B, and C) in the UCI database are randomly used as experimental datasets. The traditional K-means clustering algorithm, Clara algorithm, and the improved algorithm of this research are used for comparative analysis. The details of the experimental dataset are as follows.

As shown in Table 1, the total number of samples in the three datasets varies, and the number of attributes is also different. There are three cluster centers in dataset B and dataset C and four in dataset A. The traditional K-means clustering algorithm, Clara algorithm, and the improved algorithm of this research are run 10 times to compare the accuracy of clustering. The results are as follows.

As shown in Figure 7, in the three datasets, the accuracy of the clustering results of the traditional K-means clustering algorithm is lower than the Clara algorithm and the improved algorithm in this paper. The accuracy of the improved algorithm in this paper is higher than the traditional K-means clustering algorithm and Clara algorithm. Among them, in dataset B, the improved algorithm has the highest accuracy, up to 97.08%. In dataset C, the traditional K-means clustering algorithm has the lowest accuracy, as low as 62.17%. This shows that the improved algorithm in this

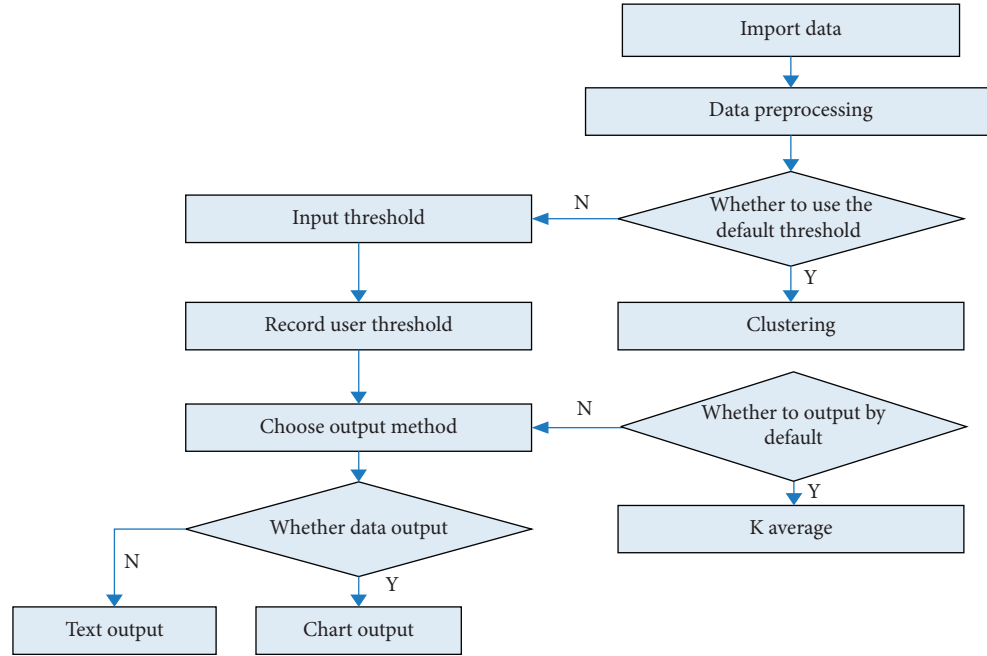


FIGURE 6: System operation flow chart.

TABLE 1: Experimental dataset.

| Dataset | Number of samples | Number of attributes | Number of clusters |
|---------|-------------------|----------------------|--------------------|
| A       | 155               | 5                    | 4                  |
| B       | 180               | 12                   | 3                  |
| C       | 540               | 9                    | 3                  |

paper has a better performance in the accuracy of the clustering results.

Then, compare the number of iterations of the three algorithms in the clustering process to analyze the stability of the algorithms. The results of the number of iterations are compared as follows.

As shown in Figure 8, the improved clustering algorithm in this paper and the traditional K-means clustering algorithm have 4 and 7 iteration times in A dataset, 6 and 11 iterations in B dataset, and 11 and 18 iterations in C dataset, respectively. This shows that the improved algorithm in this paper has strong stability in the selection of initial clustering centers.

**4.1.2. System Operation Effect.** Run a data mining system based on the improved K-means clustering algorithm of this paper to analyze 1804 customers of digital financial institutions. Use the system to divide customers into four categories: high-income customers, low-income customers, customers to be developed, and consumer customers. The criteria for dividing customers are as follows.

As shown in Table 2, there is not much difference in age between the four types of customers, mainly middle-aged people. The salient feature of high-income customers is that their annual income exceeds 100,000 and their deposits exceed one million. The salient feature of low-income

customers is that the annual income is less than 40,000 and the deposit is less than 500,000. The notable feature of the customers to be developed is that they have deposits of more than 500,000, but they have not been used for financial management or loans. The salient feature of consumer customers is that their annual income and deposits are relatively high, the amount used for financial management is higher than the annual income, and the loan amount exceeds 200,000.

After the customer classification standard is established, the type distribution of 1804 customers is analyzed, and the results are as follows.

As shown in Figure 9, among the 1804 customers of digital inclusive financial institutions, the number of customers to be developed is the largest, accounting for 46.12% of the total, and the number of high-income customers is the least, accounting for 10.75% of the total. The main types of customers who bring benefits to financial institutions are high-income customers with the smallest proportion. They often purchase wealth management products and are good at making reasonable use of the inclusive financial system to generate income for themselves.

**4.2. Evaluation and Risk Backtesting of Robo-Advisor Product Models.** Use backtesting analysis to compare the performance of robo-advisor product portfolios with performance benchmarks and evaluate the quality of product design based on the calculation of various indicators. The five risk-level investment portfolios are the capital preservation group, income group, growth and income group, growth group, and capital appreciation group. Compare the performance of the five portfolios from February 1 to 10 and the corresponding performance benchmarks. The results are as follows.

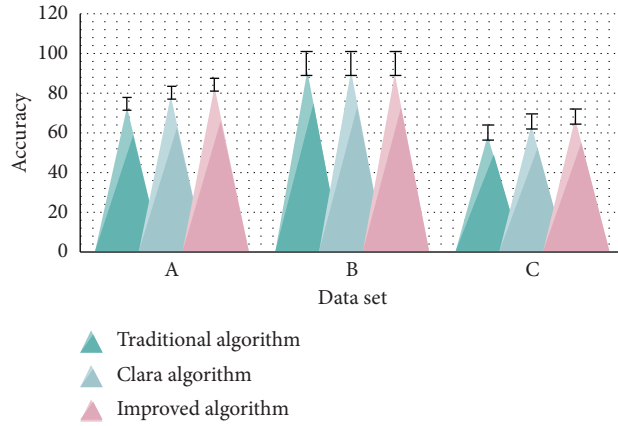


FIGURE 7: Accuracy of clustering results of different algorithms in different datasets.

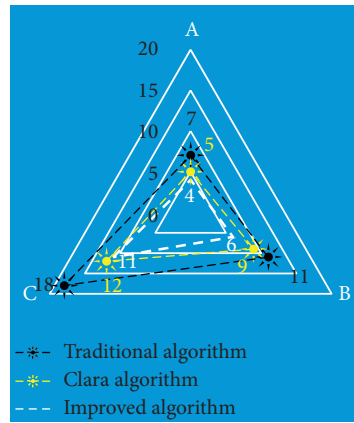


FIGURE 8: Comparison of the number of clustering iterations.

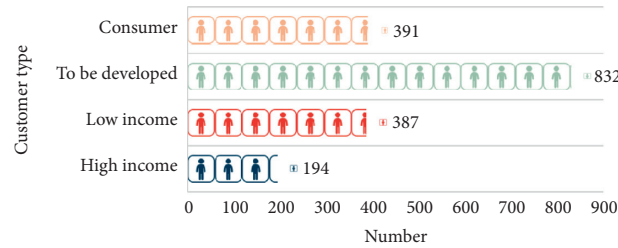


FIGURE 9: Distribution of the number of customers.

TABLE 2: Customer classification standard.

| Standard category    | High income | Low income | To be developed | Consumer |
|----------------------|-------------|------------|-----------------|----------|
| Age                  | 30–50       | <30        | 20–40           | 20–45    |
| Annual income        | >100000     | <40000     | >100000         | >100000  |
| Deposit              | >1000000    | <500000    | >50000          | >500000  |
| Financial management | >100000     | <100000    | 0               | >120000  |
| Loan amount          | >100000     | <30000     | 0               | >200000  |

As shown in Figure 10, select the RMB seven-day notice deposit interest rate, the five-year bank fixed deposit interest rate, half of the sum of the CSI All-in-One Index and the All-

in-China Securities Bond, 70% of the All-in-China Securities Index, and 30% of the All-in-China Securities Bond. The Hehe and China Securities Growth Index are used as the

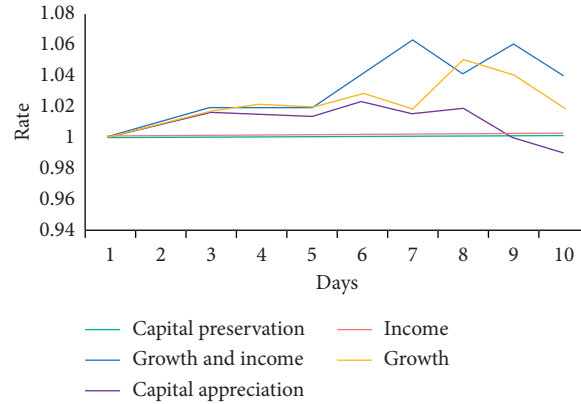


FIGURE 10: Comparison of five risk-level investment portfolios and performance benchmarks.

TABLE 3: Backtest results of the five major risk grade portfolios.

|                   | Capital preservation | Income   | Growth and income | Growth   | Capital appreciation |
|-------------------|----------------------|----------|-------------------|----------|----------------------|
| Interval return   | 0.3564%              | 0.5377%  | -0.0969%          | 2.9108%  | 5.3314%              |
| Range fluctuation | 0.1348%              | 0.3581%  | 3.1252%           | 4.5537%  | 6.7002%              |
| Maximum drawdown  | -0.3455%             | -0.7553% | -3.1124%          | -5.10295 | -7.5337%             |
| Sharpe ratio      | 1.58                 | 1.21     | -0.08             | 0.59     | 0.79                 |

TABLE 4: Comparison of capital budget before system use.

| Item number | Budget item     | Budget amount | Actual amount | Difference |
|-------------|-----------------|---------------|---------------|------------|
| H1          | Income          | 6320000       | 6200000       | 120000     |
| H2          | Tax refund      | 100000        | 120000        | -20000     |
| H3          | Cost of sales   | 4130000       | 3900000       | 230000     |
| H4          | Wage            | 1220000       | 1080000       | 140000     |
| H5          | Travel expenses | 100000        | 90000         | 10000      |
| H6          | Financial costs | 40000         | 30000         | 10000      |

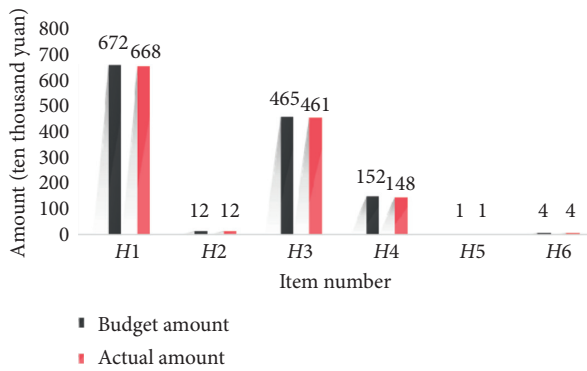


FIGURE 11: Comparison of fund budget after system use.

performance benchmarks for the five risk levels. As can be seen from the figure, the performance of the growth and income portfolio was slightly lower than the performance benchmark on February 9 and February 10 and was better than the performance benchmark in the rest of the time. Other combinations have consistently outperformed performance benchmarks during the analyzed time period.

A backtest analysis of the five major risk-level investment portfolios and the backtest results are as follows.

As shown in Table 3, the combination of growth and income is in a state of diminishing returns in the interval, and other combinations are in an increasing state in both the range of returns and fluctuations. This shows that the construction of the five risk-level investment portfolios is reasonable.

#### 4.3. Effects and Suggestions of Robo-Advisor Supervision

**4.3.1. Effect of Robo-Advisor Supervision.** Analyze the fund budget of an inclusive financial institution before and after using the data mining system of this research. The 6 items of income, tax rebates, sales costs, wages, travel expenses, and financial costs (the item numbers are H1, H2, H3, H4, H5, and H6) are budgeted and compared with the actual amount, and the results are as follows.

As shown in Table 4, before the system was used, there was a big difference between the financial institution's capital budget and actual amount. Among them, the difference in cost of sales is the largest, with a difference of 230,000, followed by wages and income, with a difference of 140,000 and 120,000, respectively. After the system is put into use, compare the fund budget and actual amount of these six projects again, and the results are as follows.

As shown in Figure 11, after the system was put into use, the difference between the financial institution's capital budget and the actual amount has narrowed a lot. The budgets for tax rebates, travel expenses, and financial costs were consistent with the actual amounts. The sales cost, which had the biggest difference before, was controlled at 40,000 yuan after the system was used. This shows that the use of the system has greatly helped financial institutions' capital budgets and can improve the accuracy of budgets.

**4.3.2. Suggestions on the Supervision of Robo-Advisors.** Based on the above analysis results, the improvement of the robo-advisory regulatory system must innovate regulatory methods, promote financial pilots, use regulatory technology, and effectively prevent risks brought by opportunities and must rely on the existing financial supervision system and laws and regulations for supervision. The synergy of self-regulatory organizations between industries must be brought into play. Financial institutions must establish and improve dispute resolution and loss compensation mechanisms and information disclosure systems.

## 5. Conclusions

Complexity science is an emerging form of research that reveals the operating rules of complex systems. Complex systems can be found in every corner of life. The development of digital inclusive finance has gone through the four stages of microfinance, microfinance, inclusive finance, and digital finance. A data mining system based on complex systems and clustering algorithms can effectively cluster, help financial institutions analyze customer information, improve the accuracy of capital budgets, and reduce risks for financial institutions.

It is high-income customers who bring benefits to inclusive financial institutions because they often buy financial products and are good at making rational use of the inclusive financial system to generate income for themselves. In this paper, the construction of the five risk-level portfolio is very reasonable. After the risk backtest, we find that except the growth and income portfolio, the performance of other portfolios is better than the performance benchmark, and the interval return and interval volatility are increasing.

We must innovate regulatory methods, promote financial pilots, use regulatory technology, and effectively prevent risks brought by opportunities and must rely on the existing financial supervision system and laws and regulations for supervision. The synergy of self-regulatory organizations between industries must be brought into play. Financial institutions should establish and improve dispute resolution and loss compensation mechanisms and information disclosure systems. Only in this way can the problems that arise in the supervision of robo-advisors be improved, and the robo-advisor system can be improved.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The author declares no conflicts of interest.

## Acknowledgments

This work was supported by the Beijing Social Science Foundation Co-Construction Project "Research on Financial Support System for Coordinated Development of Strategic Emerging Industries in Beijing-Tianjin-Hebei Region."

## References

- [1] L. Guo, "Regulating investment robo-advisors in China: problems and prospects," *European Business Organization Law Review*, vol. 21, no. 1, pp. 69–99, 2020.
- [2] M. Mompeán, A. Nogales, and D. V. Laurents, "Complex system assembly underlies a two-tiered model of highly delocalized electrons," *The Journal of Physical Chemistry Letters*, vol. 7, no. 10, pp. 1859–1864, 2016.
- [3] S. Lehuta, S. R. Mahévas, and Y. Vermard, "Reconciling complex system models and fisheries advice: practical examples and leads," *Aquatic Living Resources*, vol. 29, no. 2, p. 208, 2016.
- [4] I. K. M. Yan, X. Yi, and H. Zhang, "Digital financial inclusion and consumption smoothing in China," *China & World Economy*, vol. 28, no. 1, pp. 64–93, 2020.
- [5] D. Jung, C. V. Dorner, and H. Pusmaz, "Designing a robo-advisor for risk-averse, low-budget consumers," *Electronic Markets*, vol. 28, no. 3, pp. 367–380, 2018.
- [6] D. G. Dessavre, J. E. Ramirez-Marquez, and K. Barker, "Multidimensional approach to complex system resilience analysis," *Reliability Engineering & System Safety*, vol. 149, no. May, pp. 34–43, 2016.
- [7] C. B. Keating and P. F. Katina, "Complex system governance: concept, utility, and challenges," *Systems Research and Behavioral Science*, vol. 36, no. 5, pp. 687–705, 2019.
- [8] S. H. Bae, J. R. Cho, and W. B. Jeong, "Free and transient responses of linear complex stiffness system by Hilbert transform and convolution integral," *Smart Structures and Systems*, vol. 17, no. 5, pp. 753–771, 2016.
- [9] L. Yu and W. Wang, "An improved construction and optimization of complex system model of aerobics teaching in colleges," *Boletín Técnico/Technical Bulletin*, vol. 55, no. 15, pp. 159–165, 2017.
- [10] S. H. Wang, S. X. Zhang, and Y. Li, "Research on selective maintenance decision-making method of complex system considering imperfect maintenance," *Bingong Xuebao/Acta Armamentarii*, vol. 39, no. 6, pp. 1215–1224, 2018.
- [11] N. D. S. Dias, V. Meyer, and D. Mamedio, "Nursing practices and learning in a complex system," *European Journal of Social Sciences*, vol. 56, no. 2, pp. 149–159, 2018.
- [12] L. D. Bobo and A. S. Race, "Race as a complex adaptive system," *Du Bois Review: Social Science Research on Race*, vol. 15, no. 02, pp. 211–215, 2019.
- [13] P. F. Katina, C. B. Keating, J. C. Pyne, and J. M. Bradley, "Systemic intervention methods supporting complex system governance initiatives," *International Journal of System of Systems Engineering*, vol. 8, no. 3, p. 285, 2018.
- [14] G. O. C. Bongomin, P. Yourougou, and J. C. Munene, "Digital financial innovations in the twenty-first century: do transaction tax exemptions promote mobile money services for financial inclusion in developing countries?" *Journal of*

- Economic and Administrative Ences*, vol. 36, no. 3, pp. 185–203, 2019.
- [15] P. Traynor, J. K. Butler, and B. Reaves, “FinTechSec: addressing the security challenges of digital financial services,” *IEEE Security & Privacy*, vol. 15, no. 5, pp. 85–89, 2017.
- [16] A. Aarti, “Madhurima, factors impacting the adoption of m-banking: understanding brand India’s potential for financial inclusion,” *Journal of Asia Business Studies*, vol. 11, no. 1, pp. 22–40, 2017.
- [17] S. Naumenkova, D. S. Mishchenko, and D. Dorofeiev, “Digital financial inclusion: evidence from Ukraine,” *Investment Management and Financial Innovations*, vol. 16, no. 3, pp. 194–205, 2019.
- [18] S. D. Meena, M. Sriram, and N. Sundaram, “Digital financial inclusion is a need of the hour: an investigation amongst bank account holders in Vellore district of Tamil Nadu, India,” *International Journal of Applied Business and Economic Research*, vol. 15, no. 21, pp. 1–6, 2017.
- [19] N. . Sundaram, “Financial inclusion in India: a review,” *International Journal of Applied Engineering Research*, vol. 11, no. 3, pp. 1575–1578, 2016.
- [20] W. Ahn, H. S. Lee, H. Ryou, and K. J. Oh, “Asset allocation model for a robo-advisor using the financial market instability index and genetic algorithms,” *Sustainability*, vol. 12, no. 3, p. 849, 2020.
- [21] X. Cheng, F. Guo, J. Chen, K. Li, Y. Zhang, and P. Gao, “Exploring the trust influencing mechanism of robo-advisor service: a mixed method approach,” *Sustainability*, vol. 11, no. 18, p. 4917, 2019.
- [22] D. Gao, “Blackrock robo-advisor 4.0: when artificial intelligence replaces human discretion,” *Strategic Change*, vol. 27, no. 4, pp. 285–290, 2018.
- [23] G. Kpf, “Erste erfahrungen mit robo-advisorn-eine vergleichende analyse,” *Deutsche Molker Ztung*, vol. 139, no. 19, pp. 39–41, 2018.
- [24] D. Jung, F. V. Dorner, and S. Morana, “Robo-advisory,” *Business & Information Systems Engineering*, vol. 60, no. 1, pp. 81–86, 2018.
- [25] Z. W. Yan, “Key components and regulatory recommendations for robo-advisor in the financial services industry,” *Finance*, vol. 09, no. 6, pp. 557–563, 2019.



## Research Article

# Detection and Adaptive Video Processing of Hyperopia Scene in Sports Video

Qingjie Chen<sup>1,2</sup> and Minkai Dong<sup>3</sup> 

<sup>1</sup>School of Sports and Health, Linyi University, Linyi 276000, Shandong, China

<sup>2</sup>School of Political Science and Public Administration, Shandong University, Jinan 250100, Shandong, China

<sup>3</sup>P.E. Department, Shanghai University of Finance and Economics, Yangpu, Shanghai 200433, China

Correspondence should be addressed to Minkai Dong; [dong.minkai@mail.shufe.edu.cn](mailto:dong.minkai@mail.shufe.edu.cn)

Received 23 October 2020; Revised 17 December 2020; Accepted 29 December 2020; Published 12 January 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Qingjie Chen and Minkai Dong. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the research of motion video, the existing target detection methods are susceptible to changes in the motion video scene and cannot accurately detect the motion state of the target. Moving target detection technology is an important branch of computer vision technology. Its function is to implement real-time monitoring, real-time video capture, and detection of objects in the target area and store information that users are interested in as an important basis for exercise. This article focuses on how to efficiently perform motion detection on real-time video. By introducing the mathematical model of image processing, the traditional motion detection algorithm is improved and the improved motion detection algorithm is implemented in the system. This article combines the advantages of the widely used frame difference method, target detection algorithm, and background difference method and introduces the moving object detection method combining these two algorithms. When using Gaussian mixture model for modeling, improve the parts with differences, and keep the unmatched Gaussian distribution so that the modeling effect is similar to the actual background; the binary image is obtained through the difference between frames and the threshold, and the motion change domain is extracted through mathematical morphological filtering, and finally, the moving target is detected. The experiment proved the following: when there are more motion states, the recall rate is slightly better than that of the VIBE algorithm. It decreased about 0.05 or so, but the relative accuracy rate increased by about 0.12, and the increase ratio is significantly higher than the decrease ratio. Departments need to adopt effective target extraction methods. In order to improve the accuracy of moving target detection, this paper studies the method of background model establishment and target extraction and proposes its own improvement.

## 1. Introduction

With the maturity of computer technology, especially multimedia technology, and the processing and analysis theory of digital images, video images, as more direct and richer information carriers, are becoming more and more important research objects. In recent years, with the introduction of highway and digital earth concepts and the widespread application of the Internet, video image information has become an important source and means for humans to obtain and use information. The detection and tracking of target images is based on dynamic image analysis combined with image recognition and image tracking methods to detect targets in image sequences. The process of recognition and follow-up tracking is very important in the

field of image processing. After inspection, it was found that there was a sentence segmentation phenomenon, so the incomplete sentence was deleted directly. Inspection technology is the intersection of image processing technology, machine vision technology, and artificial intelligence technology. Therefore, detection technology has broad prospects in scientific theoretical research and also has broad prospects in practical engineering applications.

The research on video processing technology in foreign countries started early, almost accompanied by the birth of black-and-white TV sets, but limited by the technical level, the development is slow. The combination of embedded system and computer vision and image processing technology forms an object-oriented embedded video processing technology. Shukla and Sharma proposed that moving target

detection and moving target tracking algorithms are the most basic components, especially target detection algorithms, which have a direct impact on the overall performance of video surveillance systems and have been a hot spot in the field of image processing and machine vision. After inspection, it was found that there was a sentence segmentation phenomenon, so the incomplete sentence was deleted directly [1]. Yang et al. proposed an RFID security monitoring system that combines motion sensors and RFID modules. The system application relies on motion detection sensors to detect moving targets [2]. Chebi et al. used frame difference sharpness pattern matching to detect vehicles in motion and incorporated the background removal function into a small portable camera, which can better segment the foreground in the background area [3].

The domestic video surveillance industry initially developed from closed-circuit television surveillance. From simple analog video surveillance to the current purely digital network video surveillance, surveillance systems have been widely used in China recently, and their performance is comparable to that of large foreign companies. Quite a gap. Wang et al. uses the interframe difference method in the detection of moving objects in the video. Compared with the optical flow method, the advantage of the frame difference method is that the calculation speed is very fast, but there is still a big disadvantage, that is, in the detection process, the target object is easily detected [4]. Huang et al. proposed to use Wiener filtering to model and used it to predict the pixel value of the model and regarded the pixels that deviate from the estimated value as the former scenic spot [5]. Huang et al. assumed that the pixel value will change linearly with Gaussian changes over time, so a single Gaussian distribution is adopted to simulate the background model. This method mainly uses a Gaussian distribution to represent the characteristic value of a pixel and detects the pixel of the image. The eigenvalue size of will match the Gaussian distribution to classify the pixels [6].

Moving target detection and tracking are at the bottom of the vision system and are the basis for various subsequent advanced processing such as target classification and behavior understanding. This paper uses an adaptive background subtraction algorithm on a fixed single node to better complete the extraction of the moving area under a complex dynamic background and uses the static and dynamic characteristics of the moving target for multitarget tracking. The model of the tracking system under the condition that a single node can be moved is discussed, and a system model is proposed. In this paper, the contour of the target is detected, and the smallest bounding rectangle of the largest contour is used as the tracking frame of MeanShift, combined with the detection results to track the selected target, and the algorithm is applied to the UAV video to detect and track the target.

## 2. Detection and Adaptive Video Processing of Hyperopia Scene in Sports Video

*2.1. Background Modeling Based on Gaussian Mixture Model.* The  $K$ -Gaussian model is used to model every pixel in RGB color space to represent every pixel in the image. According

to the priority  $\omega_I, t (\omega_I, t = 1, I = 2k)$ , the  $K$ -Gaussian distribution used to describe the color distribution of each point has different weights from high to high  $\omega_I/\sigma_I$ . Let us say you have a low order [7, 8]. Given an appropriate background weight and threshold  $T$ , only the previous distribution within this threshold is considered as background distribution, and the other distribution is foreground distribution. At time  $t$ ,  $X_t$  is the pixel value, and the probability density function can be written as a linear combination of  $K$ -Gaussian distribution.

$$P(x_t) = \sum_{i=1}^k \frac{\omega_{i,t}}{(2\pi)^{d/2} |\text{Cov}_{t,i}|^{1/2}}, \quad (1)$$

where  $\omega_I, t, \mu_I, t, \text{Cov}_i$ , and  $t$  are the weight, mean, and covariance matrices of the  $i$  Gaussian distribution at time  $t$ , respectively. The  $K$ -Gaussian distribution is sorted in descending order of priority ( $\omega_I/\sigma_I$ ), while the previous  $b$  Gaussian distribution is used to represent the background distribution, for example,

$$B = \arg \min \left( \sum_{i=1}^b \omega_{i,t} > T \right), \quad (2)$$

where  $t$  is the set background threshold. For the convenience of calculation, it is assumed that the red, green, and blue components of pixels are independent of each other.

$$\begin{aligned} x_t &= (x_{r,t}^t, x_{g,t}^t, x_{b,t}^t), \\ \mu_{j,t} &= (\mu_{j,t}^r, \mu_{j,t}^g, \mu_{j,t}^b), \\ \text{Cov}_{i,t} &= (\text{Cov}_{i,t}^r, \text{Cov}_{i,t}^g, \text{Cov}_{i,t}^b). \end{aligned} \quad (3)$$

*2.2. Automatic Extraction of Stadium Main Area Based on Histogram Statistics.* The document used statistical information on the differences of color elements to distinguish legal areas. For the main characteristics of the green soccer field and the lawn tennis court, the extraction algorithm is used based on the difference in component color information. In the lawn area, the green element of the pixel is larger than the red and blue elements. The differences between the green component and the other two components were calculated and then the color characteristics of the lawn were obtained by threshold treatment. The grass area is extracted from the image, and the binary image is achieved by threshold processing [9, 10].

$$I(x, y) = \begin{cases} 1, & G(x, y) - R(x, y) > T \text{ and } (G(x, y) - B(x, y)) > T_b, \\ 0, & \text{other.} \end{cases} \quad (4)$$

Among them,  $R(x, y)$ ,  $G(x, y)$ , and  $B(x, y)$  are pixel values of color image and  $T_R$  and  $T_B$  are two thresholds which can be adjusted according to the position. Choosing reasonable parameters in the green football field, the extraction accuracy of the main areas of the stadium is relatively accurate [11, 12]. However, this algorithm can only be

used in lawns and other places with prior knowledge and cannot be extracted adaptively.

After inspection, it was found that there was a sentence segmentation phenomenon, so the incomplete sentence was deleted directly. Due to the influence of noise and other factors in the imaging process, there are many small noises or small areas in the main area of the segmented stadium, so it is impossible to remove them directly by removing small areas. According to the nature of noise, the method of eliminating block motion noise can solve this problem well [13, 14]. The blocking operation used in this paper is noise algorithm which is as follows: the image is divided into  $n \times n$  blocks, and the statistical value of stadium color in each block is greater than a certain percentage; the block is regarded as stadium; otherwise, the stadium block is considered. The formula is

$$B(i, j) \begin{cases} 1, & \text{if } \sum_{y=N*j}^{y=N*(j+1)} \sum_{x=N*i}^{x=N*(i+1)} G(x, y) > T \times (N^2), \\ 0, & \text{other.} \end{cases} \quad (5)$$

Among them,  $B(I, J)$  is a binary result graph composed of  $n \times n$  blocks,  $G(x, y)$  is the primary extracted main area of the course, and  $t$  is the proportion of the area. In most cases,  $t$  is 0.6 and  $n$  is 36.

It is assumed that the gray value of the pixel at the same point does not change significantly in two consecutive frames, so it can be considered to be approximately equal as shown in the following formula:

$$f(x + dx, y + dy, t + dt) = f(x, y, t). \quad (6)$$

Among them,  $f(x, y, t)$  is the gray scale of pixel  $(x, y)$  at time  $t$  and  $f(x + dx, y + dy, t + dt)$  is the gray scale of  $(x, y)$  moving to pixel  $(x + dx, y + dy)$ . Since the foundation of the algorithm is based on gray level consistency, it is also necessary to assume that formula (7) holds, namely:

$$\frac{\partial f(x + dx, y + dy, t + dt)}{\partial t} = 0. \quad (7)$$

After the second-order Taylor formula expansion and simplification of formula (7), we can obtain the optical flow field calculation formula:

$$\frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} + \frac{\partial f}{\partial t} = 0. \quad (8)$$

Among them,  $(\partial f / \partial x)$  can be understood as the gradient of the brightness of the image in the horizontal direction,  $(dx / dt)$  can be understood as the gradient in the vertical direction,  $(\partial f / \partial t)$  can be understood as the gradient in time, and  $(dx / dt)$  and  $(dy / dt)$  are the optical flow in the horizontal direction and optical flow in the vertical direction, respectively.

**2.3. Eigen Image Filtering.** Among various image processing methods, adjusting pixels is the most basic image processing method. It can adjust the changes of the image under different illuminations through the logarithmic conversion of the pixel value. In addition, rendering is also a kind of image processing, and reasonable rendering can make the image

clearer [15, 16]. Recently, Finlayson proposed a new intrinsic image algorithm, and by calibrating the camera's sensitive equipment, the color image is converted into one-dimensional gray-scale image, thus eliminating the influence of shadow. We use this method in this article because there are few restrictions on this method, and it is not very strict. At the same time, this paper provides a method to improve the calibration technology and expand the scope of the algorithm.

Assuming that the Lambert model can be used in the imaging process of the camera photoreceptor, an integral formula can be used to describe the response of each color channel of the photoreceptor [17, 18].

$$\rho_k(x, y) = \sigma(x, y) \int E(\lambda, x, y) S(\lambda, x, y) Q_k(\lambda) d\lambda, \quad (9)$$

where  $k = B_{GR}$  represents the index of the color channel,  $Q_k$  is the spectral response function of the  $k$ th color sensing device,  $\sigma(x, y)$  is the Lambert light dark coefficient, which is the dot product of the normal vector of the reflection plane and the direction vector of the incident light,  $E(\lambda, x, y)$  is the spectral energy distribution function of the incident light, and  $S(\lambda, x, y)$  is the reflection function of the object surface. If the following two conditions are met at each point of the photosensitive device: the spectral response function of each color channel is a Dirac function and the illuminance function of the incident light can use Planck model, then  $Q_k(\lambda)$  and  $E(\lambda, x, y)$  can be expressed as follows:

$$Q_k(\lambda) = q_k \delta(\lambda - \lambda_k), \quad (10)$$

$$E(\lambda, T) \approx I_c \lambda^{-5} e^{-(c_2/T\lambda)}, \quad (11)$$

where  $Q_k$  is the response strength of the inductor. Because every pixel needs to be considered, the pixel position parameter in the function is removed.  $I$  refers to the absolute parameter of light control  $c_1$  and  $c_2$  refer to the intensity of light, and,  $t$  refers to the absolute temperature. Substituting formulas (7) and (8) into formula (6) yields

$$P_k = \sigma I_{c_1} \lambda_k^{-5} e^{-(c_2/T\lambda)} S(\lambda_k) q_k. \quad (12)$$

Now establish the ratio of the two channels:

$$r_k = \frac{P_k}{P_p}. \quad (13)$$

If  $p = G$ , then  $k = R, B$ ; take the natural logarithm for formula (10) and use two values to form a two-dimensional vector ( $R/G$ ):

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \log\left(\frac{S_R}{S_G}\right) \\ \log\left(\frac{S_B}{S_G}\right) \end{bmatrix} + \frac{1}{T} \begin{bmatrix} e_R - e_G \\ e_B - e_G \end{bmatrix}, \quad (14)$$

where  $\lambda_1 = R/G$  and  $\lambda_2 = B/k$ ; from (11), it can be seen that the first term on the right side of the formula is a constant, whose value is determined jointly by the reflecting surface

and the camera, and the second term is determined by the incident light temperature and the camera. With the change of the temperature of the incident light, the image points of the same medium are distributed along a straight line in the two-dimensional plane, and this straight line is only related to the camera, not to the incident light and the reflection plane [19, 20].

The basic idea of the interframe difference method is to obtain the shape of the moving target by performing the difference operation on two consecutive frames of the video image sequence.

The change between the images  $f_k(x, y)$  and  $f_{k+1}(x, y)$  of frame  $k$  and frame  $k + 1$  is represented by a two-finger differential image  $D(x, y)$ , as shown in the following formula:

$$D(x, y) = \begin{cases} 1, & \text{if } |f_{k+1}(x, y) - f_k(x, y)| > T, \\ 0, & \text{others.} \end{cases} \quad (15)$$

In formula (15), 0 corresponds to the area where the two frames of images have not changed and 1 corresponds to the area that has changed. The flowchart of the interframe difference method is shown in Figure 1.

**2.4. Background Recognition Modeling.** The common methods of background model building and updating are the statistical average method, coefficient updating method, Gauss model method, etc [21, 22]. The complexity of various methods of background model building and updating is different, and the effect of background model obtained is also different.

The statistical average method is to average the continuous video sequence and take the average value as the pixel value in the current background model. Assuming that the current video sequence is the  $n$ th frame, then the current time background model is common, and the formula is as follows [23, 24]:

$$B_{n-1}(i, j) = \frac{\sum_{k=1}^{n-1} f_k(i, j)}{n-1},$$

$$B_{n-1}(i, j) = (1 - \alpha)B_{n-2}(i, j) + \alpha f_n(i, j),$$

$$\mu_{n-1}(i, j) = \frac{1}{n-1} \sum_{k=1}^{n-1} f_k(i, j), \quad (16)$$

$$\sigma_{n-1}^2(i, j) = \frac{1}{n-1} \sum_{k=1}^{n-1} [f_k(i, j) - \mu(i, j)]^2,$$

where  $f_k(i, j)$  the pixel gray value at the  $k$  image coordinate  $(i, j)$ . When the  $H$ -th image is acquired [25, 26], the system will get the gray value. According to the above formula transformation process, the improved Gaussian function formula is as follows:

$$\mu_{n-1}(i, j) = (1 - \alpha)\mu_{n-1}(i, j) + \alpha f_n(i, j),$$

$$\sigma_n^2(i, j) = (1 - \alpha)\sigma_{n-1}^2(i, j) + \alpha [f_n(i, j) - \mu_n(i, j)]^2. \quad (17)$$

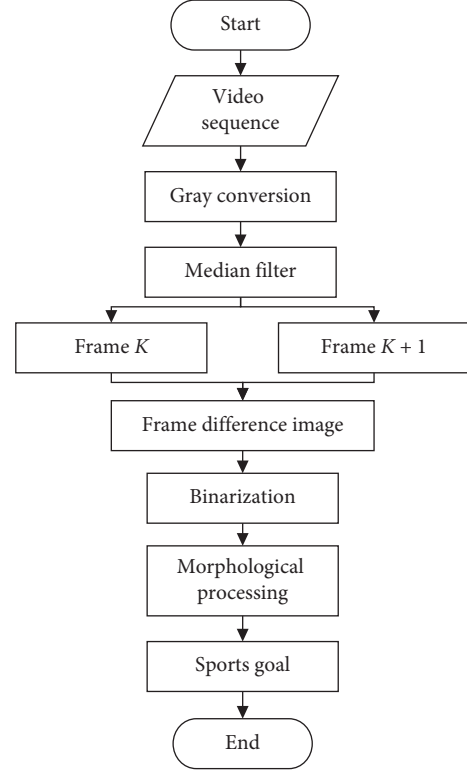


FIGURE 1: Interframe difference method flowchart.

Firstly, the first image of video sequence is used as the initial background, and the current tilt image  $f_n$  at  $n$  time in the post  $w$  sequence matrix is used with the previous moment background image  $F_0$  [27, 28]. A background model updating template is obtained by differential operation, and the common  $CB_{n-1}$  is obtained by binarization of the updated template with decision threshold  $BT_n$ :

$$BT_n(i, j) = \begin{cases} 0 \\ 1 |F_n(i, j) - CB_{n-1}(i, j)| > D. \end{cases} \quad (18)$$

Background image  $CB_n$  is updated as follows:

$$CB_n(i, j) = \begin{cases} CB_{n-1}(i, j), \\ \omega_1 CB_{n-1}(i, j) + \omega_2 F_n(i, j). \end{cases} \quad (19)$$

### 3. Hyperopia Scene Detection and Adaptive Video Processing Experiment Design

**3.1. System Design.** The image acquisition subsystem divides the video information collected by the USB camera into two parts. The first part is transmitted to the real-time monitoring client through the network; the other part enters the moving target detection subsystem. When a moving target passes in the surveillance area, the system will quickly detect the target. The subsystem calls the alarm subsystem to perform alarm work and at the same time compresses and saves the video pictures and several continuous screenshots when moving objects appear in the monitoring area through the image compression subsystem and the storage subsystem

and saves them locally. When a user wants to find a moving target, he can view the video or picture through the WEB subsystem of the network to quickly find the target object. The process is shown in Figure 2.

**3.2. Test Subject.** In order to compare the effect and processing speed of the remote scene detection system in this article, the system in this article is compared with several common systems that use other algorithms. The main comparison algorithms are GMG algorithm, GMM detection algorithm, IMBS detection algorithm, KDE detection algorithm, and VIBE detection algorithm; the algorithm in this article is an improved VIBE detection algorithm. The video images of the experimental database are used for qualitative comparison, and these experimental data are collected for data analysis to draw conclusions. In this paper, the precision rate and recall rate are used as indicators to evaluate the detection algorithm, and the processing speed is also used as the evaluation indicator. After inspection, it was found that there was a sentence segmentation phenomenon, so the incomplete sentence was deleted directly. The larger the value, the better the detection of the algorithm; the smaller the value of processing speed, the better the detection of the algorithm.

**3.3. Experimental Method.** First, it is necessary to fill the image with holes to remove small target impurities, that is, when there are disturbances in the background, such as the shaking of leaves, or the misdetection caused by changes in illumination, this will cause some false small targets in the foreground image. And there will be some defects and holes in the detected moving target, which will have an impact on the subsequent contour detection and tracking. In intelligent monitoring, some false targets will cause false alarms and incomplete targets will cause missed detection. Secondly, the circumscribed rectangle of the detected target after improvement needs to be used as the initial target of the MeanShift tracking algorithm. The contour of the final target image will be detected. After the moving target is obtained, the moving target to be tracked is selected, that is, select tracking target area, tracking the target.

**3.4. Statistical Data Processing Method.** SPSS23.0 software was used for data processing, and the count data were expressed in percentage (%);  $k$  is the number of data in this experiment,  $\sigma^2$  is the variance of all survey results, and  $P < 0.05$  indicates that the difference is statistically significant. The formula for calculating reliability is as follows:

$$a = \frac{k}{k-1} \left( 1 - \frac{\sum \sigma_i^2}{\sigma^2} \right). \quad (20)$$

## 4. Experimental Hyperopia Scene Detection and Adaptive Video Processing

**4.1. Evaluation Index System Based on Index Reliability Testing.** Reliability refers to the stability and reliability of the questionnaire. This article adopts the  $\alpha$  coefficient method created by L.J. Cronbach. The  $\alpha$  coefficient can be obtained

by reliability analysis in SPSS software. It is generally believed that the  $\alpha$  coefficient above 0.8 indicates that the effect of the index setting is very good, and above 0.7 is also acceptable. Here we analyze the reliability of each type of object, and the reliability index we choose for each type of object is slightly different. The results are shown in Table 1.

It can be seen from Table 1 that there are certain differences in the processing results of different scenes, but these processing results can be optimized technically, so the difference between the processing effects of different scenes has an acceptable impact on this experiment ( $\alpha > 0.7$ ). In order to better illustrate the detection effect, corresponding indicators are used to evaluate the detection effect. According to the segmented real foreground image, many indicators are used to evaluate the background modeling algorithm.

### 4.2. Comparison and Analysis of Detection Algorithms

**4.2.1. Effect Analysis of Baseline Video Library Scene.** First, we analyze the scenes of the baseline video library and compare the effect analysis of the algorithm used in the system with that of other algorithms. Here, five scenes are selected for comparative analysis: highway scene, indirect motion scene, scene with similar color, small target scene, and hyperopia scene. The results of the target scene and the hyperopia scene are shown in Table 2. We make a bar graph based on this result, as shown in Figure 3.

It can be seen from Figure 3 that although the GMG detection algorithm can detect a relatively complete target, the algorithm is more sensitive to illumination changes and background disturbance; the GMM detection algorithm is more obvious in the detection of the target wheel, but it is not easy to find it inside the target and when the target is hidden. When the background is more complex, the detection effect is not very good; the IMBS detection algorithm has a good detection effect, but it is very sensitive to light changes and background disturbance; the KDE detection algorithm has a bad detection effect, and the algorithm is more sensitive to the contour of the object. The detection rate is very high; the VIBE detection algorithm is more robust to illumination and background disturbance; the detection algorithm in this paper maintains the advantages of the VIBE algorithm and further improves the detection accuracy.

**4.2.2. Effect Analysis of Dynamic Background Video Library Scene.** We analyze the scenes of the dynamic background video library and compare the effect analysis of the algorithm used in the system with that of other algorithms. Here, five scenes are selected for comparative analysis: the oar shaking scene, water wave shaking scene, leaf shaking scene, fountain scene, and hyperopia scene. The results of the fountain scene and the hyperopia scene are shown in Table 3. We make a histogram based on this result, as shown in Figure 4.

It can be seen from Figure 4 that the GMG detection algorithm is very sensitive to background disturbances, and background disturbances have a great influence on the detection effect; the detection effect of GMM detection algorithm is not ideal for background disturbances; the IMBS

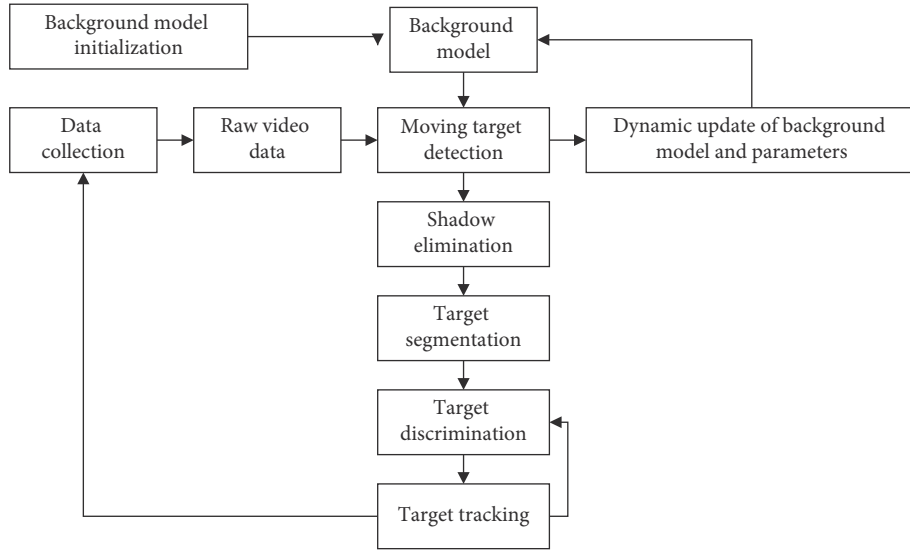


FIGURE 2: Moving target detection and tracking program structure diagram.

TABLE 1: Summary of reliability test results.

| Category                   | Index combination   | Alpha coefficient ( $\alpha$ ) |
|----------------------------|---------------------|--------------------------------|
| Highway scene              | Heuristic rule      | 0.8632                         |
|                            | Spatial correlation |                                |
| Indirect motion scene      | Heuristic rule      | 0.8447                         |
|                            | Spatial correlation |                                |
| Scenes with similar colors | Heuristic rule      | 0.7365                         |
|                            | Spatial correlation |                                |
| Small target scene         | Heuristic rule      | 0.7746                         |
|                            | Spatial correlation |                                |
| Hyperopia scene            | Heuristic rule      | 0.8124                         |
|                            | Spatial correlation |                                |

detection algorithm is very sensitive to the interference of leaf shaking, and there is a lot of false detection rate; VIBE detection algorithm still has a little false detection; the detection algorithm in this paper reduces the target of false detection for the rippling of water waves and the shaking of leaves, and it can also detect moving targets.

#### 4.2.3. Effect Analysis of Camera Shake Video Library Scene.

We analyze the scenes of the camera shake video library and compare the effect analysis of the algorithm used in the system with that of other algorithms. Here, five scenes are selected for comparative analysis, namely: badminton court scene, highway scene, zebra crossing scene, road scene, and hyperopia scene. The results are shown in Table 4, and we make a line chart based on this result, as shown in Figure 5.

It can be seen from Figure 5 that the GMG detection algorithm has a relatively high false detection rate for camera shake. Although the complexity of the GMM detection algorithm is reduced, its detection accuracy is not very high; although the KDE detection algorithm has a high detection rate, there are still errors, while the IMBS detection algorithm has large errors, and its detection effect needs to be further improved; the false detection rate of the VIBE

detection algorithm is relatively reduced, but there will still be some false warnings. The detection algorithm in this article guarantees the detection accuracy and reduces the false detection rate.

### 4.3. Evaluation Index Analysis

**4.3.1. Analysis of the Accuracy Index of Dynamic Background Video Library Scene Detection.** We analyze the detection accuracy index of dynamic background video library scenes and compare the algorithm used in this paper with other algorithms to analyze the motion state of people in the video and determine its accuracy. Here, five scenes are selected for comparative analysis. They are badminton court scene, basketball court scene, street scene, park scene, and hyperopia scene. The results are shown in Table 5. We make a combined picture based on this result, as shown in Figure 6.

It can be seen from Figure 6 that for the five scenes with dynamic background, there are fast-moving people or indirect-moving people, there are changes in illumination, and the leaves are slightly shaking. The improved algorithm in this paper is relatively stable in the detection accuracy index value. Compared with the original VIBE algorithm, it has increased to a certain percentage, so it verifies the effectiveness of the improved method in this paper.

**4.3.2. Analysis of the Recall Rate Index of Dynamic Background Video Library Scene Detection.** We analyze the detection recall rate index for the scene of the dynamic background video library and compare the algorithm used in this paper with other algorithms to analyze the motion state of people in the video and determine the recall rate. The results are shown in Table 6. This result makes a bar graph, as shown in Figure 7.

The improved algorithm in this paper has different recall rates for different scenarios. As can be seen from Figure 7, the recall rate is higher than that of the VIBE algorithm when

TABLE 2: Baseline video library scene effect analysis data.

| Algorithm                           | Highway scene | Indirect motion scene | Scenes with similar colors | Small target scene | Hyperopia scene |
|-------------------------------------|---------------|-----------------------|----------------------------|--------------------|-----------------|
| GMG detection algorithm             | 1.42          | 1.6                   | 1.83                       | 2.39               | 2.56            |
| GMM detection algorithm             | 1.62          | 1.85                  | 2.33                       | 2.45               | 3.11            |
| IMBS detection algorithm            | 1.89          | 2.07                  | 2.74                       | 2.81               | 3.4             |
| KDE detection algorithm             | 2.17          | 2.73                  | 3.09                       | 3.22               | 3.58            |
| VIBE detection algorithm            | 2.76          | 2.8                   | 3.16                       | 3.52               | 3.71            |
| Detection algorithm of this article | 2.84          | 3.35                  | 3.66                       | 3.89               | 4.99            |

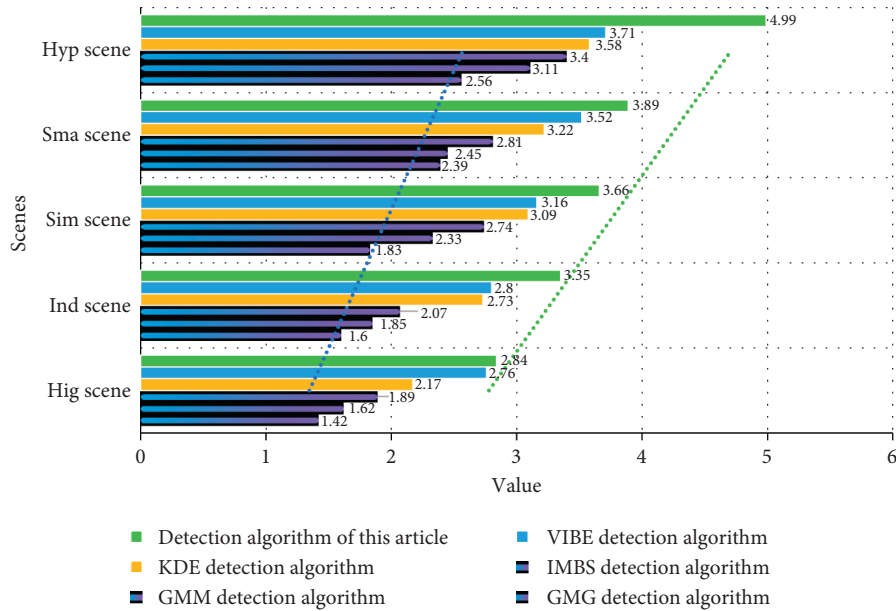


FIGURE 3: Effect analysis diagram of baseline video library scene.

TABLE 3: Effect analysis data of dynamic background video library scene.

| Algorithm                           | Oar shaking scene | Water wave shaking scene | Leaf shaking scene | Fountain scene | Hyperopia scene |
|-------------------------------------|-------------------|--------------------------|--------------------|----------------|-----------------|
| GMG detection algorithm             | 1.58              | 1.35                     | 1.68               | 2.32           | 2.71            |
| GMM detection algorithm             | 2.11              | 1.57                     | 2.08               | 2.33           | 3.00            |
| IMBS detection algorithm            | 2.21              | 2.04                     | 2.13               | 2.81           | 3.31            |
| KDE detection algorithm             | 2.39              | 2.17                     | 3.11               | 3.01           | 3.52            |
| VIBE detection algorithm            | 3.18              | 2.42                     | 3.04               | 3.77           | 3.79            |
| Detection algorithm of this article | 3.23              | 3.12                     | 3.63               | 3.93           | 4.16            |

the motion state is relatively uniform, and when there are more motion states, the recall rate is slightly better than that of the VIBE algorithm. It decreased about 0.05 or so, but the relative accuracy rate increased by about 0.12, and the increase ratio is significantly higher than the decrease ratio.

**4.3.3. Analysis of the Speed Index of the Dynamic Background Video Library Scene Detection.** We analyze the detection speed index of the scene of the dynamic background video library and compare the algorithm used in this paper with other algorithms to analyze the motion state of people in the video and determine the detection speed. The results are shown in Table 7. The result is a histogram, as shown in Figure 8.

The improved algorithm in this paper has different detection speed changes for different scenarios. From Figure 8, it can be seen that when the motion state is relatively uniform, the increase in detection speed is not too large. When there are more motion states, the detection speed increases. It can be clearly observed.

**4.3.4. Analysis of the Accuracy Index of Camera Shake Video Library Scene Detection.** We analyze the detection accuracy index of the camera shake video library scene and compare the algorithm used in this paper with other algorithms to analyze the motion state of people in the video and determine its accuracy. Here, five scenes are selected for

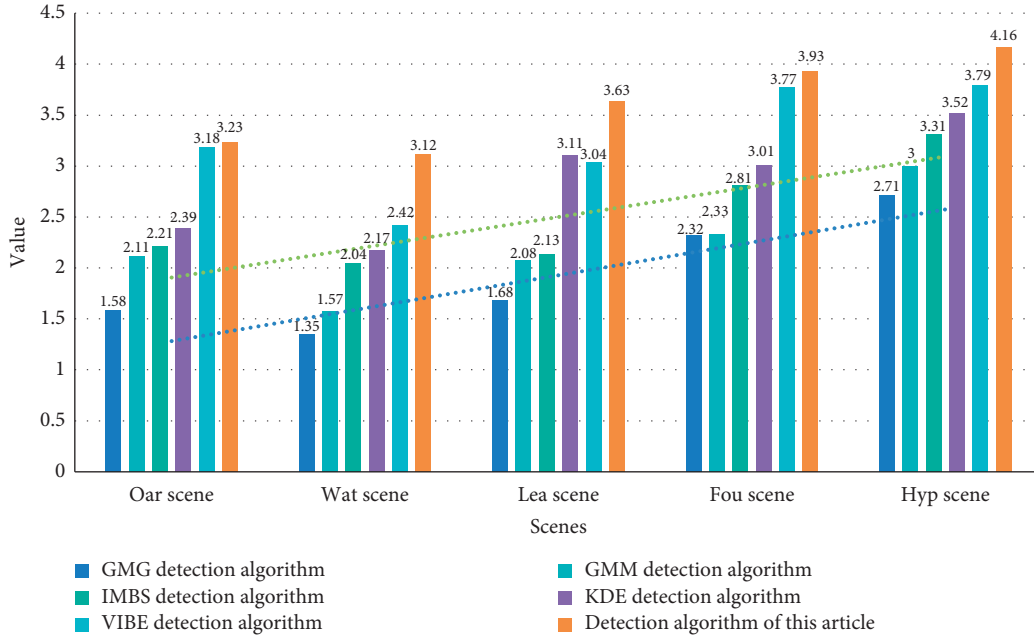


FIGURE 4: Effect analysis diagram of dynamic background video library scene.

TABLE 4: Effect analysis data of camera shake video library scene.

| Algorithm                           | Badminton court scene | Highway scene | Zebra crossing scene | Road scene | Hyperopia scene |
|-------------------------------------|-----------------------|---------------|----------------------|------------|-----------------|
| GMG detection algorithm             | 5.01                  | 4.12          | 3.84                 | 3.45       | 4.87            |
| GMM detection algorithm             | 1.95                  | 2.19          | 3.02                 | 5.52       | 4.32            |
| IMBS detection algorithm            | 2.50                  | 1.32          | 1.88                 | 2.59       | 3.70            |
| KDE detection algorithm             | 2.81                  | 4.15          | 6.20                 | 5.65       | 6.53            |
| VIBE detection algorithm            | 3.41                  | 3.12          | 4.83                 | 5.19       | 7.03            |
| Detection algorithm of this article | 5.12                  | 4.91          | 4.80                 | 4.82       | 7.79            |

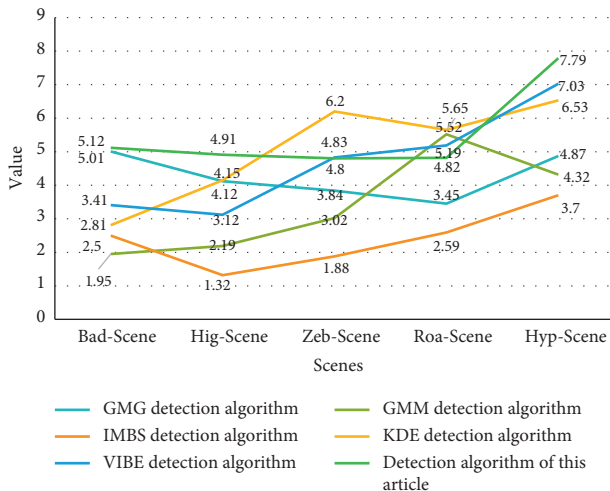


FIGURE 5: Effect analysis diagram of camera shake video library scene.

comparison and analysis. These are subway station scene, highway scene, zebra crossing scene, road scene, and far vision scene. The results are shown in Table 8. We make a histogram based on this result, as shown in Figure 9.

It can be seen from Figure 9 and from the experimental data that for the five complex scenes of camera shake, the effect of the improved algorithm in this paper is relatively stable in the detection accuracy index value, and the false detection rate is lower than that of other algorithms. The VIBE algorithm has been improved, so for camera shake, the detection effect is better than that of other algorithms.

*4.3.5. Analysis of the Recall Rate Index of Camera Shake Video Library Scene Detection.* We analyze the detection recall rate index of the camera shake video library scene and compare the algorithm used in this paper with other algorithms to analyze the motion state of people in the video and judge the recall rate. The results are shown in Table 9. This result makes a histogram, as shown in Figure 10.

The improved algorithm in this article has different recall rates for different scenarios. As shown in Figure 10, when the motion state is relatively uniform, the recall rate is higher than that of the VIBE algorithm, and when the motion state increases, it is higher than that of the VIBE algorithm. There is only an increase of about 0.02, but the relative accuracy rate has increased by about 0.08, and the improvement ratio is significantly higher.



TABLE 5: Data of accuracy index for dynamic background video library scene detection.

| Detection accuracy     | GMG  | GMM  | IMBS | KDE  | VIBE | This article |
|------------------------|------|------|------|------|------|--------------|
| Badminton court scene  | 0.96 | 0.35 | 0.97 | 0.83 | 0.93 | 0.98         |
| Basketball court scene | 0.94 | 0.24 | 0.88 | 0.72 | 0.95 | 1.00         |
| Street scene           | 0.96 | 0.60 | 0.97 | 1.00 | 0.95 | 1.00         |
| Park scene             | 0.89 | 0.40 | 0.87 | 0.63 | 0.86 | 0.91         |
| Hyperopia scene        | 0.80 | 0.97 | 0.68 | 0.72 | 0.83 | 0.95         |

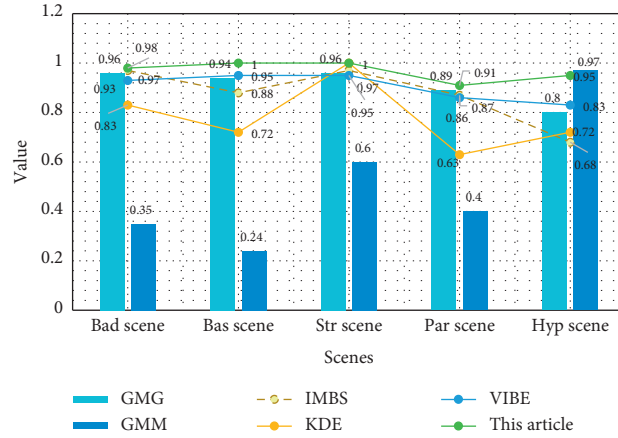


FIGURE 6: The accuracy index map of scene detection for dynamic background video library.

TABLE 6: Data of recall rate index for dynamic background video library scene detection.

| Test recall rate       | GMG  | GMM  | IMBS | KDE  | VIBE | This article |
|------------------------|------|------|------|------|------|--------------|
| Badminton court scene  | 0.92 | 0.40 | 0.92 | 0.65 | 0.85 | 0.89         |
| Basketball court scene | 0.74 | 0.24 | 0.75 | 0.47 | 0.82 | 0.85         |
| Street scene           | 0.84 | 0.60 | 0.83 | 0.90 | 0.80 | 0.82         |
| Park scene             | 0.66 | 0.30 | 0.69 | 0.23 | 0.62 | 0.67         |
| Hyperopia scene        | 0.79 | 0.69 | 0.59 | 0.63 | 0.73 | 0.85         |

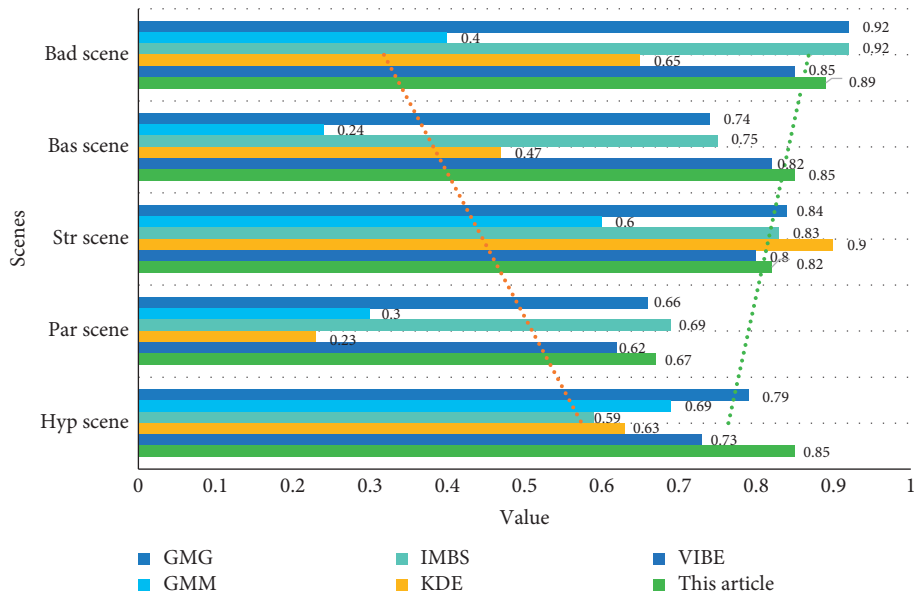


FIGURE 7: Recall rate indicator diagram for scene detection of dynamic background video library.

TABLE 7: Data of speed indicators for scene detection of dynamic background video library.

| Detection speed        | GMG  | GMM  | IMBS | KDE  | VIBE | This article |
|------------------------|------|------|------|------|------|--------------|
| Badminton court scene  | 0.09 | 0.12 | 0.07 | 0.16 | 0.05 | 0.01         |
| Basketball court scene | 0.16 | 0.17 | 0.09 | 0.21 | 0.13 | 0.06         |
| Street scene           | 1.03 | 0.99 | 0.97 | 1.07 | 0.89 | 0.79         |
| Park scene             | 0.86 | 0.79 | 1.03 | 0.94 | 0.82 | 0.66         |
| Hyperopia scene        | 0.17 | 0.26 | 0.19 | 0.31 | 0.22 | 0.13         |

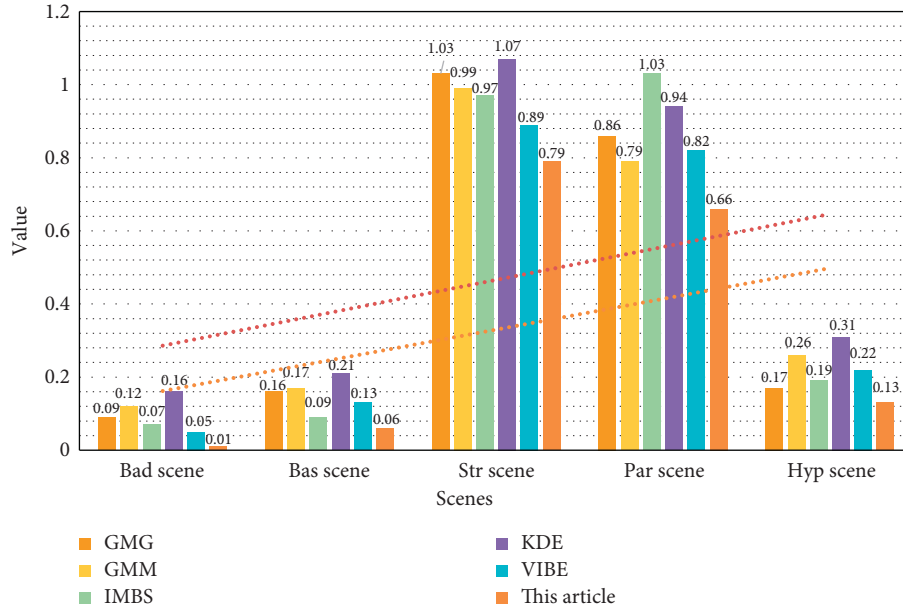


FIGURE 8: Speed detection speed indicator diagram for scene of dynamic background video library.

TABLE 8: Data of accuracy index for camera shake video library scene detection.

| Detection accuracy   | GMG  | GMM  | IMBS | KDE  | VIBE | This article |
|----------------------|------|------|------|------|------|--------------|
| Subway station scene | 0.80 | 0.60 | 0.97 | 0.94 | 0.98 | 0.98         |
| Highway scene        | 0.57 | 0.93 | 0.85 | 0.95 | 0.95 | 0.95         |
| Zebra crossing scene | 0.85 | 0.30 | 0.85 | 0.80 | 0.94 | 0.97         |
| Road scene           | 0.82 | 0.92 | 0.90 | 0.95 | 0.95 | 0.97         |
| Hyperopia scene      | 0.43 | 0.73 | 0.89 | 0.76 | 0.83 | 0.94         |

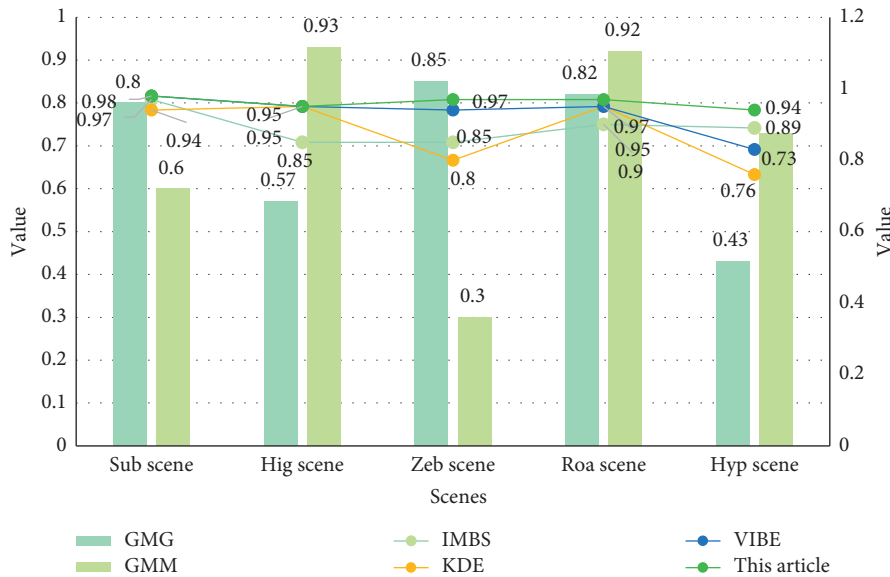


FIGURE 9: The accuracy index chart of camera shake video library scene detection.

TABLE 9: Data of recall rate index for camera shake video library scene detection.

| Detection accuracy   | GMG  | GMM  | IMBS | KDE  | VIBE | This article |
|----------------------|------|------|------|------|------|--------------|
| Subway station scene | 0.31 | 0.48 | 0.88 | 0.44 | 0.73 | 0.86         |
| Highway scene        | 0.23 | 0.63 | 0.75 | 0.77 | 0.78 | 0.85         |
| Zebra crossing scene | 0.53 | 0.43 | 0.75 | 0.33 | 0.62 | 0.89         |
| Road scene           | 0.40 | 0.68 | 0.90 | 0.82 | 0.91 | 0.93         |
| Hyperopia scene      | 0.47 | 0.77 | 0.54 | 0.73 | 0.82 | 0.90         |

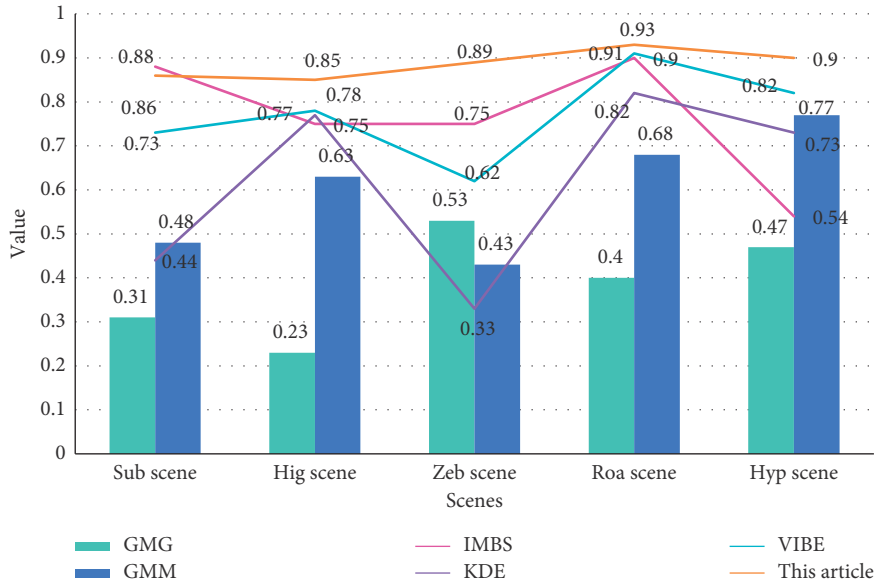


FIGURE 10: Recall rate index graph for camera shake video library scene detection.

TABLE 10: Data of speed indicators for camera shake video library scene detection.

| Detection accuracy   | GMG  | GMM  | IMBS | KDE  | VIBE | This article |
|----------------------|------|------|------|------|------|--------------|
| Subway station scene | 0.89 | 0.76 | 1.03 | 0.88 | 0.64 | 0.37         |
| Highway scene        | 0.36 | 0.13 | 0.19 | 0.28 | 0.15 | 0.01         |
| Zebra crossing scene | 0.35 | 0.23 | 0.27 | 0.65 | 0.33 | 0.06         |
| Road scene           | 0.16 | 0.34 | 0.26 | 0.22 | 0.19 | 0.01         |
| Hyperopia scene      | 0.17 | 0.26 | 0.19 | 0.31 | 0.22 | 0.13         |

4.3.6. *Analysis of the Speed Index of Camera Shake Video Library Scene Detection.* We analyze the detection speed index of the camera shake video library scene and compare the algorithm used in this paper with other algorithms to analyze the motion state of people in the video and determine the detection speed. The results are shown in Table 10. The result is a histogram, as shown in Figure 11.

The improved algorithm in this paper has different detection speeds for different scenarios. From Figure 11, it can be seen that when the motion state is relatively uniform, the detection speed will increase slightly due to the increase in the number of people to be detected. When the motion state increases, it will not increase because of the increase in the number of people tested.

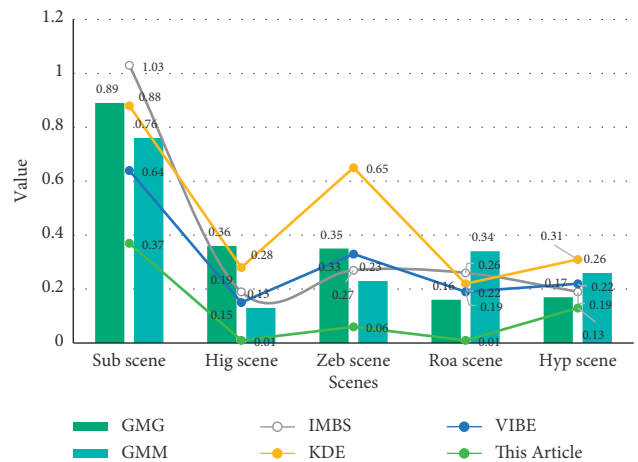


FIGURE 11: Speed detection indicators for camera shake video library scene.

## 5. Conclusions

This topic is derived from a target image search project of a certain research. It detects intrusive moving targets in the monitored warning area. The background of the video is more complicated. Since the characteristics of the moving target are unknown, the background subtraction method is used. The key step of this method is the establishment of the background

model, the target extraction, and the update of the background model method; only the generated background model has good adaptability to the changes of the complex background and can reflect the real background image, and the target needs to be extracted after the background model is established. Departments need to adopt effective target extraction methods. In order to improve the accuracy of moving target detection, this paper studies the method of background model establishment and target extraction and proposes its own improvement.

Afterwards, an improved detection algorithm is used to clean and delete the shadow area of the moving target in order to better track the target. Combining the improved VIBE algorithm and the shadow removal of multifeature fusion, it has been verified by experiments to effectively remove the shadow of the detected target, making the detected moving target more accurate. Finally, we detect the maximum contour of the target from which the shadow is removed, use the minimum bounding rectangle of the target as the initial window of MeanShift tracking, and combine the detection results to track the target. Through comparison and verification, the MeanShift tracking algorithm combined with the improved detection algorithm in this paper has better tracking effect than the MeanShift algorithm without combined improvement. In this paper, the improved detection and tracking methods are applied to the video images captured by the drone, and the moving targets in the video are detected and the targets are tracked. It can be seen from the experimental results that the moving targets in the video taken by the drone are clearer. The target detection and tracking effect is relatively ideal, which further proves the effectiveness of the improved method in this paper.

By applying the detection and tracking system of moving objects to sports video, the system will process the collected motion information data and obtain the motion parameters of human body. It is very important to send some data to improve the quality of athletes and coaches and expand the development direction of athletes. This paper is based on the background of the research and development of sports video object detection and tracking system, through the Gauss model to separate the scene, through the eigen extraction to extract the characteristics of the players in the stadium, and finally through the difference method to identify the video content, which opens a new idea for the detection and adaptive processing of the athletes in sports video.

### Data Availability

No data were used to support this study.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Acknowledgments

This study was supported by the National Social Science Fund (Research on the development strategy of urban and rural mass sports from the perspective of the concept of "coordination") (17BTY084).


### References

- [1] D. Shukla and M. Sharma, "A novel video scene change detection using successive estimation of statistical measure and HiBiSLI method," *International Journal of Information Technology*, vol. 11, no. 1, pp. 47–54, 2019.
- [2] X. Yang, D. Wang, W. Zhang, P. Peng, and L. Feng, "Detection of ventricular tachycardia and fibrillation using adaptive variational mode decomposition and boosted-CART classifier," *Biomedical Signal Processing and Control*, vol. 39, pp. 219–229, 2018.
- [3] H. Chebi, M. D. Acheli, and M. Kesraoui, "Dynamic detection of abnormalities in video analysis of crowd behavior with DBSCAN and neural networks," *Advances in Science, Technology and Engineering Systems Journal*, vol. 1, no. 5, pp. 56–63, 2016.
- [4] Y. Wang, A. W.-C. Y. Hu, and C.-T. Li, "ENF based video forgery detection algorithm," *International Journal of Digital Crime and Forensics*, vol. 12, no. 1, pp. 131–156, 2020.
- [5] Z. Huang, J. B. Sui, and G. Jiang, "An intelligent ship image/video detection and classification method with improved regressive deep convolutional neural network," *Complexity*, vol. 2020, Article ID 1520872, 11 pages, 2020.
- [6] C. C. Huang, C. E. Lee, and V. L. L. Thing, "A novel video forgery detection model based on triangular polarity feature classification," *International Journal of Digital Crime and Forensics*, vol. 12, no. 1, pp. 14–34, 2020.
- [7] Q. Abbas, M. A. Ibrahim, and M. A. Jaffar, "Video scene analysis: an overview and challenges on deep learning algorithms," *Multimedia Tools and Applications*, vol. 77, no. 16, pp. 20415–20453, 2018.
- [8] Z. Shao, Z. L. Wang, and W. W. Wu Du, "Saliency-aware convolution neural network for ship detection in surveillance video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 3, pp. 781–794, 2020.
- [9] L. Pan, T. Pezzuti, and C. W. Pechmann, "Hyperopia and frugality: different motivational drivers and yet similar effects on consumer spending," *Journal of Business Research*, vol. 95, pp. 347–356, 2019.
- [10] S. Feldman, M. M. W. Peterseim, R. H. Trivedi et al., "Detecting high hyperopia: the plus lens test and the spot vision screener," *Journal of Pediatric Ophthalmology & Strabismus*, vol. 54, no. 3, pp. 1–5, 2016.
- [11] G. S. Ying and G. E. Quinn, "Uncorrected hyperopia and preschool early literacy: results of the Vision in Preschoolers-Hyperopia in Preschoolers (VIP-HIP) Study," *Ophthalmology*, vol. 123, no. 4, pp. 681–689, 2016.
- [12] M. Garcia-Gonzalez, M. L. Iglesias-Iglesias, P. D. Rodriguez-Casanova et al., "Femtosecond laser-assisted LASIK with and without the adjuvant use of mitomycin C to correct hyperopia," *Journal of Refractive Surgery (Thorofare, N.J.: 1995)*, vol. 34, no. 1, pp. 23–28, 2018.
- [13] C. Roesler and T. Kohnen, "Changes of functional optical zone after LASIK for hyperopia and hyperopic astigmatism," *Journal of Refractive Surgery (Thorofare, N.J.: 1995)*, vol. 34, no. 7, pp. 476–481, 2018.
- [14] M. D. Basar, A. D. Duru, and A. Akan, "Emotional state detection based on common spatial patterns of EEG," *Signal, Image and Video Processing*, vol. 14, no. 3, pp. 473–481, 2020.
- [15] W. Sun, J. D. Yan, and C. Sun, "Small-scale moving target detection in aerial image by deep inverse reinforcement learning," *Soft Computing*, vol. 24, no. 8, pp. 5897–5908, 2020.
- [16] Naoki, Tanimura, Natsuko et al., "Visual function and functional decline in patients with waterclefts," *Investigative*

- Ophthalmology & Visual Science*, vol. 60, no. 10, pp. 3652–3658, 2019.
- [17] Y. Lu and S. An, “Research on sports video detection technology motion 3D reconstruction based on hidden Markov model,” *Cluster Computing*, vol. 23, no. 3, pp. 1899–1909, 2020.
  - [18] M. Alizadeh and M. Sharifkhani, “Compressed domain moving object detection based on CRF,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 3, pp. 674–684, 2020.
  - [19] T. Rice, “Hudl video software developer: agile sports technologies,” *International Sport Coaching Journal*, vol. 3, no. 3, pp. 364–365, 2016.
  - [20] Y. Yuan, J. Chu, L. Leng et al., “A scale-adaptive object-tracking algorithm with occlusion detection,” *Eurasip Journal on Image & Video Processing*, vol. 2020, no. 1, pp. 1–15, 2020.
  - [21] M. Qi, A. Y. Wang, and J. Luo, “Sports video captioning via attentive motion representation and group relationship modeling,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 8, pp. 2617–2633, 2020.
  - [22] A. Tejero-De-Pablos, T. Y. Nakashima, and M. N. LinnaYokoya, “Summarization of user-generated sports video by using deep action recognition features,” *IEEE Transactions on Multimedia*, vol. 20, no. 8, pp. 2000–2011, 2018.
  - [23] K. C. Rahtu and A. Bhandari, “A euclidean distance based super resolution method for sub pixel target detection in hyper spectral images,” vol. 15, no. 2, pp. 104–124, 2020.
  - [24] H. Wei and G. Bu, “The Fusion algorithm based on frame difference and optical flow for sports video detection,” *C e Ca*, vol. 42, no. 4, pp. 1498–1502, 2017.
  - [25] A. Khan, D. Windridge, and J. Kittler, “Multilevel Chinese takeaway process and label-based processes for rule induction in the context of automated sports video annotation,” *IEEE Transactions on Cybernetics*, vol. 44, no. 10, pp. 1910–1923, 2017.
  - [26] Z. Yu, F. Shuo, S. Xiaohua et al., “Research on tracking algorithm for fast-moving target in sport video,” *Journal of Computational & Theoretical Nanoence*, vol. 14, no. 1, pp. 230–236, 2017.
  - [27] M. Zhuang, “Sports video structure analysis and feature extraction in long jump video,” *International Journal of Multimedia and Ubiquitous Engineering*, vol. 9, no. 9, pp. 83–92, 2016.
  - [28] W. . Xinghua, “Application research on sports video analysis using object segmentation algorithms,” *International Journal of Simulation: Systems, Ence & Technology*, vol. 17, no. 36, pp. 8.1–8.9, 2016.

## Research Article

# Study on Mechanical Characteristics of Energy-Absorbing and Anti-Scour Bolts

Zhi Tang <sup>1,2</sup>, Hao Wu,<sup>1</sup> Jinguo Lv,<sup>1</sup> Zhuangzhuang Xin,<sup>1</sup> and Wenbo Zuo<sup>1</sup>

<sup>1</sup>School of Mechanics and Engineering, Liaoning Technical University, Fuxin 123000, China

<sup>2</sup>Civil and Environmental Engineering School, University of Science and Technology Beijing, Beijing 100083, China

Correspondence should be addressed to Zhi Tang; tangzhi0127@163.com

Received 12 August 2020; Revised 10 October 2020; Accepted 24 December 2020; Published 8 January 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Zhi Tang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to improve the impact resistance and mechanical performance of anchor rods and satisfy the requirements for supporting rockburst roadways, the energy balance equation of the energy-absorbing support and roadway surrounding rock system is established. Moreover, to effectively prevent rockburst disasters, the energy criterion for roadway instability is derived. From the perspective of an energy-absorbing support, a yield-absorbing anti-shock anchor composed of a rod body, tray, constant resistance energy-absorbing device, and special-shaped nut is designed and developed; compared with ordinary anchor rods, this rod has stronger mechanical properties for resisting impact. Theoretical and numerical simulation studies show that the energy-absorbing device has a repeatable deformation failure mode and a constant yield force. The paper also presents the principle involved in the design of anti-shock bolt supports. The energy-absorbing support not only effectively guides and controls the release and conversion of impact energy but also consumes the impact energy in the buffering process of the anchor to ensure the stability of surrounding rock and support protection system. This study aims to provide reference for roadway support design and to improve rock bolts used in rockburst roadways.

## 1. Introduction

The current energy system requires coal to further perform its function as “ballast stone.” Coal is therefore anticipated to continue as China’s leading energy source in the future. The continuous large-scale exploitation of resources in previous decades, however, has exhausted the country’s shallow coal resources. Future exploitations of coal resources are expected to reach depths of 1000–2000 m. At depths beyond 1000 m, the geological conditions of coal resources become more complex, the original rock stress increases, the surrounding rock deformation of roadways becomes severe, and ground pressure impact becomes the primary threat leading to disaster in many mining areas [1–5]. In Poland, from 1949 to 1982, a total of 3097 large-scale destructive shocks occurred, causing 401 deaths and damaging 1, 20, 000 m of wells and roads. From 1910 to 1978, the Ruhr Coal Mine in Germany experienced 283 large-scale destructive shocks [6], and from 1933 to 2019, China experienced 9020

destructive coal mine rockbursts [7]. It is predicted that as mining depths increase, rockburst and its resulting series of accidents will increasingly become severe, extremely retarding the exploitation of deep coal resources. The control of surrounding rock of deep roadways is one of the key problems in deep mining theory. In coal mining, the inadequacy of roadway support is the main problem encountered in roadway surrounding rock control as well as rockburst prevention and control.

For rockburst roadway support, a new type of suction bolt with a constant working resistance and good elongation capacity has been developed [8–11]. In reality, suction bolts have been studied by scholars from various countries for over 20 years. In the early 1990s, Ortlepp [12] proposed the concept of an energy-absorbing support system, and Jager [13] developed the first true suction bolt—the cone bolt. It was not until the late 1990s that resin-anchored tapered bolts were introduced [14]. In recent years, having gained in-depth knowledge on energy-

absorbing bolt and supports, experts have developed a variety of supports such as Garford anchor, Durabar anchor, yielding Secura anchor, Roofex anchor, and anchor conceived by Varden et al. [15–17]. With the focus set on the actual surrounding rock deformation of coal mine roadway sustaining ground pressure impact, the negative Poisson's ratio material of the anchor rod (cable) is investigated, and a high constant resistance large deformation anchor cable is designed. Most existing anchor rods are not designed based on attaining strong impact resistance; consequently, they are weak against impact. From the perspective of impact resistant design, most investigations focus on the research and development of new materials and new types of support members. Although the application of these research results has resulted in certain headways, it has also considerably increased the cost of roadway support. In view of the limitations of existing bolts and the dearth of special requirements for tunnel support in sustaining ground pressure impact, a new type of energy-absorbing anti-scour bolt is designed and developed. In order to provide reference for the design and improvement of rock bolt support in rockburst roadway, the structure composition and working principle of this new bolt are investigated.

The specific contributions of this paper include the following:

- (1) The energy balance equation of the energy-absorbing support and roadway surrounding rock system is established based on the principle of energy conservation.
- (2) The energy criterion for the instability of roadway supported by an energy-absorbing system is deduced. Accordingly, the foregoing provides a theoretical basis for energy-absorbing support and anti-scour design.
- (3) This paper presents the principle involved in energy absorption and anti-scour.
- (4) In view of the advantages afforded by energy-absorbing supports, a type of energy-absorbing anti-scour bolt, composed of rod body, tray, constant resistance energy-absorbing device, and special nut, is designed and developed. The test results indicate that the energy-absorbing anti-scour bolt has stronger impact mechanical properties than the ordinary bolt.

The rest of this paper is organized as follows. In Section 2, the energy balance equation of the energy-absorbing support and roadway surrounding rock system is established based on the principle of energy conservation. Moreover, the energy criterion for the instability of roadway supported by an energy-absorbing system is deduced. A type of energy-absorbing anti-scour bolt, composed of rod body, tray, constant resistance energy-absorbing device, and special nut, is designed in Section 3. Section 4 shows the simulation experimental results, and Section 5 concludes the paper with summary.

## 2. Energy Criterion for Instability of Roadway

Rockburst is defined as the sudden and instantaneous release of energy from the coal and rock mass around the roadway. This energy mainly emanates from a source and is transmitted to the roadway in addition to the energy accumulated through the elastic deformation of roadway surrounding rock.

The initial energy of an earthquake (as impact source) is reduced by the propagation of waves through the surrounding rock until the roadway surrounding rock is reached. Let it be assumed that the initial energy of the rockburst source is  $E_z$ , the radius of the circular roadway is  $r$ , the distance from the rockburst source to the center of the roadway is  $d$ , and the energy attenuation index of the rockburst shock wave propagating through the coal rock medium is  $\eta$ . With the foregoing, the energy ( $E_{zh}$ ) released from the impact ground pressure of the impact source is calculated as follows:

$$E_{zh} = E_z (d - r)^{-\eta}. \quad (1)$$

The elastic deformation energy of the roadway surrounding rock that has accumulated in the original rock stress field is expressed by the following equation:

$$E_3 = \iiint_V \left[ \frac{[\sigma_1^2 + \sigma_2^2 + \sigma_3^2 - 2\nu(\sigma_1\sigma_2 + \sigma_1\sigma_3 + \sigma_2\sigma_3)]}{2E} \right] dx dy dz. \quad (2)$$

Upon the initiation of roadway surrounding rock failure, the stress state of the coal rock mass rapidly changes from a three-way to a two-way force and then immediately transforms to a one-way force. According to the minimum energy principle for energy transfer, the energy required for coal rock mass failure is the failure energy under unidirectional stress, i.e., the minimum failure energy ( $E_{\min}$ ).

$$E_{\min} = \frac{\sigma_c^2}{2E}, \quad (3)$$

$$\text{or } E_{\min} = \frac{\tau_c^2}{2G}.$$

In the roadway excavation process, various forms of energy dissipation (e.g., plastic deformation of rock mass, relative joint plane slippage in surrounding rock mass, viscous flow deformation, secondary crack, and crack propagation at the crack tip of surrounding rock mass) occur. If it is presumed that the right side of the equation is  $E_q$ , then the rockburst energy ( $E_p$ ) derived from the elastic deformation energy of the surrounding rock is as follows:

$$E_p = E_3 - E_{\min} - E_q. \quad (4)$$

According to equations (1) and (4), the energy released from coal and roadway surrounding rock mass is obtained when rockburst occurs.

$$E_{cj} = E_{zh} + E_p = E_z (d - r)^{-\eta} + E_3 - E_{\min} - E_q. \quad (5)$$

When rockburst occurs, a part of the energy released by the coal rock mass around the roadway is absorbed by the roadway support structure, and the other part is released to the roadway in the form of coal rock kinetic energy.

If it is assumed that  $\delta$  denotes the maximum deformation of the roadway supporting structure, which can withstand an impact load,  $F_{zf}$ , under the surrounding rock pressure, then the energy absorbed by the elastic deformation of the roadway supporting structure can be calculated as follows:

$$E_{zf} = \frac{\delta F_{zf}}{2}. \quad (6)$$

If it is assumed that the energy released to the roadway in the form of coal rock kinetic energy is  $E_k$  when rockburst occurs, then according to the law of conservation of energy,

$$E_{cj} = E_{zf} + E_k. \quad (7)$$

That is,

$$E_z(d-r)^{-\eta} + E_3 - E_{\min} - E_q = \frac{\delta F_{zf}}{2} + E_k. \quad (8)$$

To prevent the roadway from impacting the ground pressure, the energy ( $E_k$ ) that should be released to the roadway in the form of coal rock kinetic energy must be equal to zero. If  $E_k$  is greater than zero, then

$$E_z(d-r)^{-\eta} + E_3 - E_{\min} - E_q > \frac{\delta F_{zf}}{2}. \quad (9)$$

The foregoing shows that the roadway support structure cannot fully absorb the energy released by rockburst, and the residual energy ( $E_k$ ) will be released to the roadway in the form of coal rock kinetic energy, resulting in the damage of the support structure and collapse of the surrounding rock. Equation (9) is therefore the energy criterion for roadway impact failure.

If it is assumed that the support has an energy absorption function and the energy absorbed by the energy-absorbing device is  $E_x$ , then according to equation (8), the energy conservation equation for energy absorption and roadway anti-scour support is expressed as follows:

$$E_z(d-r)^{-\eta} + E_3 - E_{\min} - E_q = \frac{\delta F_{zf}}{2} + E_x + E_r + E_k. \quad (10)$$

If the energy ( $E_{cj}$ ) released by rockburst is greater than the energy ( $E_x$ ) absorbed by the energy-absorbing device and the energy ( $F_{zf}$ ) absorbed by the elastic deformation of the support structure, then

$$E_{cj} = E_z(d-r)^{-\eta} + E_3 - E_{\min} - E_q > E_x + E_r + E_{zf}. \quad (11)$$

This shows that the support cannot completely absorb the impact pressure to release energy, and the remaining energy ( $E_k$ ) is either projected as kinetic energy or transformed into a large displacement of the surrounding rock to free space; this causes support damage and collapse of the roadway surrounding rock. Equation (11) is therefore the

energy criterion for the instability of roadway supported by an energy-absorbing system.

According to the foregoing criterion, the adoption of an energy-absorbing support in the roadway can considerably enhance the anti-impact performance of the support system as well as effectively preclude and control the occurrence of rockburst. Ultimately, the criterion provides a theoretical basis for the anti-impact design of an energy-absorbing bolt support.

### 3. Design of Energy-Absorbing Anti-Impact Anchor

**3.1. Structural Design of Energy-Absorbing Anti-Impact Anchor.** The incorporation of a constant resistance energy-absorbing device to the conventional anchor rod is adopted for the design of the energy-absorbing anti-scour anchor rod shown in Figure 1. This rod includes four parts: rod body (1), tray (2), constant resistance energy-absorbing device (3), and special-shaped nut (4). The right end of the rod body is fabricated with external threads that can be connected to the nut, and the center of the tray has a mounting hole for the rod body. The constant resistance energy-absorbing device, which has a 150–200 mm thin-walled circular tube structure, is placed between the tray and special-shaped nut. The special nut (Figure 2) is fabricated by integrating a round table block, a ring block, and a square block.

**3.2. Theoretical Analysis of Mechanical Properties of Energy-Absorbing Devices.** During impact, existing energy-absorbing devices have large load fluctuations, which do not satisfy the mechanical performance requirements for anti-scour bolt; hence, it is necessary to design a new device [18–20]. The constant resistance energy-absorbing device is designed as a thin-walled circular tube that slides along the round table block of the special nut to achieve a constant resistance yield. Let the outer diameter, inner diameter, and wall thickness of the thin-walled circular tube be  $2R_0$ ,  $2r_0$ , and  $t$ , respectively (note that the inner diameter ( $2r_0$ ) is larger and smaller than the diameters of the left and right circles of the special-shaped nut round block, respectively). Through expansion, these outer and inner diameters are made to transition to the larger outer and inner diameters, denoted as  $2R_1$  and  $2r_1$ , of the rear thin-walled tube, respectively. The round nut angle of the special-shaped nut is  $\alpha$ , the axial compressive stress is  $\sigma_z$ , the radial compressive stress is  $\sigma_n$ , the hoop tensile stress is  $\sigma_\theta$ , and the friction coefficient between the inner wall of the thin-walled tube and special-shaped nut is  $\mu$ . To analyze the expansion force on the thin-walled circular tube, a circular ring-shaped element is obtained from the expansion area bounded by two planes normal to the axis of the tube, as shown in Figure 3 [21].

$$\begin{aligned} & (\sigma_z + d\sigma_z)\pi[(R+dR)^2 - (r+dR)^2] - \sigma_z\pi(R^2 - r^2) \\ & + 2\sigma_n\pi r dR + 2\frac{\mu\sigma_n\pi r}{\tan\alpha} dR = 0. \end{aligned} \quad (12)$$



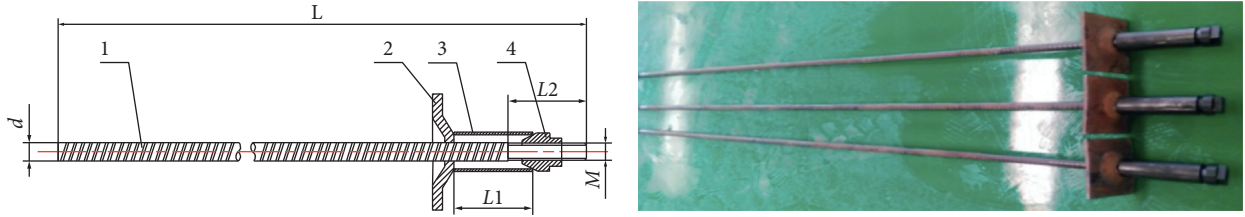


FIGURE 1: Energy-absorbing anti-impact anchor.

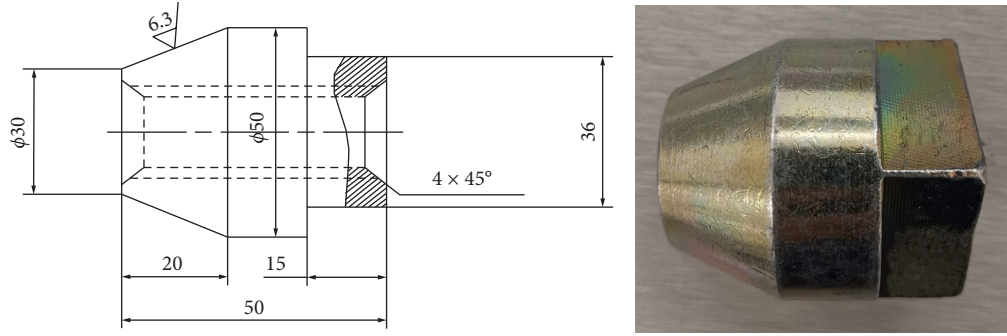


FIGURE 2: Diagram of special-shaped nut.

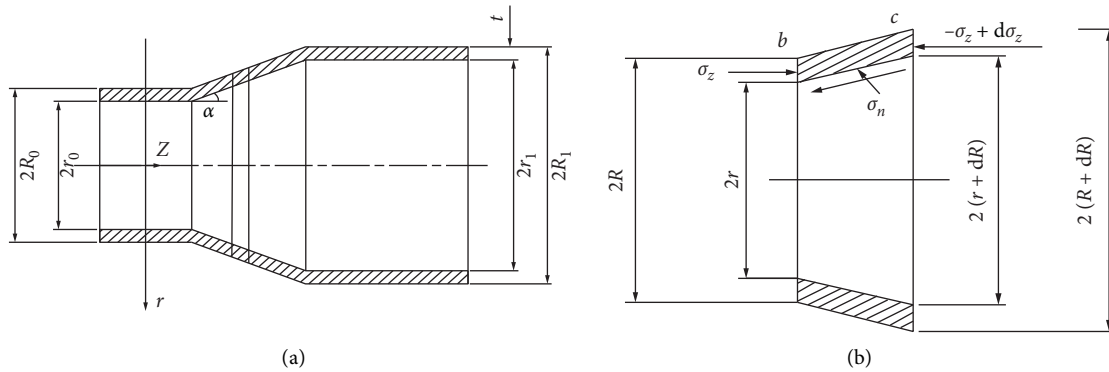


FIGURE 3: Diameter and base body. (a) Radius expansion diagram. (b) Primitive body force diagram.

Set the basic element in the  $Z$ -axis direction of the equilibrium equation.

Omitting the high-order trace yields the following:

$$(R^2 - r^2)d\sigma_z + 2[(R - r)\sigma_z + \sigma_n r (1 + \mu \cot \alpha)]dR = 0. \quad (13)$$

Assuming that  $\sigma_r$  is the combined stress of the radial compressive and frictional stresses of the toroidal base body, the following relationship can be introduced:

$$\sigma_n = \frac{\sigma_r}{1 - \mu \tan \alpha}. \quad (14)$$

If it is assumed that the wall thickness of the thin-walled tube remains constant during radius expansion, then the corresponding yield condition is

$$\sigma_z + \sigma_\theta = K, \quad (15)$$

where  $K$  is the plane deformation resistance, given as  $K = 1.155\sigma_s$ , and  $\sigma_s$  is the yield stress of the material.

For the stress analysis, half of the circular ring element is considered as shown in Figure 4.

The vertical balance equation is as follows:

$$2\sigma_\theta t = \int_0^\pi \sigma_r r d\theta \cdot \sin \theta. \quad (16)$$

Equation (15) is rewritten as equation (16).

$$\sigma_r = \frac{R - r}{r} (1.15\sigma_s - \sigma_z). \quad (17)$$

Combining equations (17) and (14) and substituting the results into equation (12) yield

$$\frac{d\sigma_z}{\sigma_z a - 1.15\sigma_s(a + 1)} = \frac{2dR}{R + r}, \quad (18)$$

where  $a = (1 + \mu \cot \alpha / 1 - \mu \tan \alpha) - 1$ .

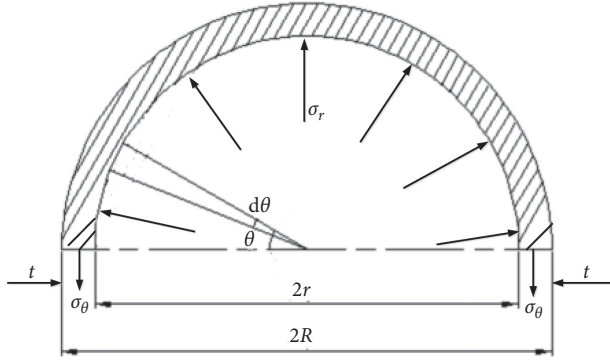


FIGURE 4: Half ring body force diagram.

When the integral of equation (18) is substituted into the boundary condition ( $R = R_0$ ,  $r = r_0$ , and  $\sigma_z = 0$ ), the yield stress of the constant energy-resistive device is given by the following:

$$\sigma_z = \beta \sigma_s \frac{a+1}{a} \left[ 1 - \left( \frac{R_1 + r_1}{R_0 + r_0} \right)^{2a} \right]. \quad (19)$$

In the foregoing,  $\beta = (K/\sigma_s)$ .

The yield force of the constant resistance energy-absorbing device can be obtained from equation (19), as follows:

$$F = \pi(R_0^2 - r_0^2)\sigma_z. \quad (20)$$

That is,

$$F_{\max} = 1.15\pi(R_0^2 - r_0^2)\sigma_s \frac{a+1}{a} \left[ 1 - \left( \frac{R_1 + r_1}{R_0 + r_0} \right)^{2a} \right]. \quad (21)$$

Based on equation (20), it can be concluded that the bearing capacity of the constant resistance energy-absorbing device is influenced by inner diameter, wall thickness, material yield strength, friction coefficient, and other parameters.

The energy absorbed can be calculated as

$$E = F\delta = 1.15\pi(R_0^2 - r_0^2)\sigma_s \frac{a+1}{a} \left[ 1 - \left( \frac{R_1 + r_1}{R_0 + r_0} \right)^{2a} \right] \left( L - \frac{r_1 - r_0}{\sin \theta} \right), \quad (22)$$

where  $L$  and  $\delta$  are the length and compression distance of the thin-walled tube, respectively.

**3.3. Numerical Simulation of Mechanical Properties of Constant Resistance Energy-Absorbing Device.** The model is established using ABAQUS finite element analysis software [22–28]. The special-shaped nut is assumed to be a 160 mm high rigid body with a 110 mm lower end diameter, a 100 mm upper end diameter, and a 10 mm high round table block. Considering the influence of the mechanical properties of thin-walled circular tube materials on energy absorption, three materials, specifically, T700L, Q550, and Q235, are simulated. The material parameters are shown in Tables 1 and 2. The thin-walled circular tube is 3 mm thick

and 150 mm long with a 105 mm inner diameter. The circular tube network is mainly a quadrilateral element with a 2 mm characteristic length. With all degrees of freedom of the special-shaped nut constrained, the thin-walled tube is placed above the nut, and a rigid plate is pressed axially from above the thin-walled tube at a constant speed of 1 m/s. An automatic point-to-surface contact mechanism is used between the rigid plate and circular tube. During the deformation process, the internal and external surfaces of the component are set to self-contact with a 0.3 friction coefficient.

The compression deformation process of the constant resistance energy-absorbing device is shown in Figure 5, and the force and energy absorption displacement curves are shown in Figures 6 and 7, respectively.

- (1) The various materials of the constant resistance energy absorbing-device have stable and repeatable deformation and failure modes during compression.
- (2) The constant resistance energy-absorbing device composed of different materials has a constant force during the yielding process, and the absorbed energy increases approximately linearly with the increase in compression distance. The yielding force and energy absorption results of the device during compression deformation are summarized in Table 1. (a) The yielding force and absorbed energy of T700L are 344 kN and 46.90 kJ, respectively, and the theoretical yielding force and absorbed energy are 347 kN and 45.47 kJ, respectively; the relative errors are 1% and 3%, respectively. (b) The simulated yielding force and absorbed energy of Q550 are 276 kN and 35.33 kJ, respectively, and the theoretical yielding force and absorbed energy are 262 kN and 35.87 kJ, respectively; the relative errors are 5% and 1%, respectively. (c) The simulated yielding force and absorbed energy of Q235 are 156 kN and 20.33 kJ, respectively, and the theoretical yielding force and absorbed energy are 112 kN and 15.10 kJ, respectively; the relative errors are 28% and 26%, respectively.

**3.4. Principles of Energy Absorption, Anti-Scour, and Anti-Scour Bolt Support.** For the energy-absorbing device to effectively function, its displacement force should be greater than the yielding force of the bar body; moreover, the breaking force of the bar body should be low. When the bearing capacity of the bolt is less than its yielding force, the energy-absorbing device will not yield to deformation. The principle of the support afforded by energy-absorbing anti-shock bolts is mainly reflected in two aspects: resistance and yield. The energy-absorbing device is not deformed and destroyed by the resisting bodies under the static pressure of the surrounding rock. The device is allowed to absorb energy to reduce impact when the ground exerts pressure. These processes are reflected in three aspects: first, the energy-absorbing device directly dissipates the impact energy of the surrounding rock

TABLE 1: Dimensions and anti-shock characteristics of thin-walled round pipes.

| Component number | Material model | Wall thickness (mm) | Diameter (mm) | Length (mm) | Simulated $F_{max}$ (kN) | Theoretical $F_{max}$ (kN) | Relative error (%) | Theoretical $E$ (kJ) | Simulated $E$ (kJ) | Relative error (%) |
|------------------|----------------|---------------------|---------------|-------------|--------------------------|----------------------------|--------------------|----------------------|--------------------|--------------------|
| 1                | T700L          | 3                   | 105           | 150         | 344                      | 347                        | 1                  | 45.47                | 46.90              | 3                  |
| 2                | Q550           | 3                   | 105           | 150         | 276                      | 262                        | 5                  | 35.87                | 35.33              | 1                  |
| 3                | Q235           | 3                   | 105           | 150         | 156                      | 112                        | 28                 | 20.33                | 15.10              | 26                 |

TABLE 2: Material parameters.

| Material model | $\text{kg}\cdot\text{m}^{-3}$ | $E$ (GPa) | $\nu$ | $\sigma_s$ (MPa) | $\sigma_b$ (MPa) | $\delta$ (%) |
|----------------|-------------------------------|-----------|-------|------------------|------------------|--------------|
| T700L          | 7850                          | 210       | 0.3   | 720              | 790              | 16           |
| Q550           | 7850                          | 200       | 0.3   | 550              | 620              | 16           |
| Q235           | 7850                          | 200       | 0.3   | 235              | 380              | 16           |

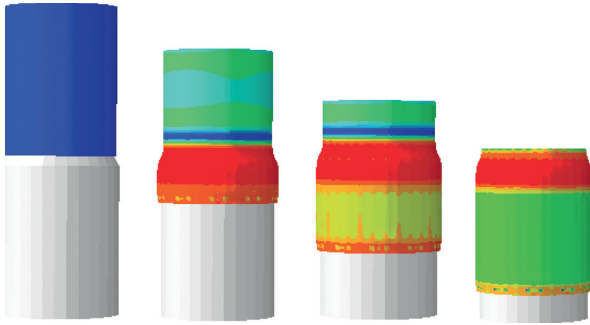


FIGURE 5: Deformation process of constant resistance energy-absorbing device.

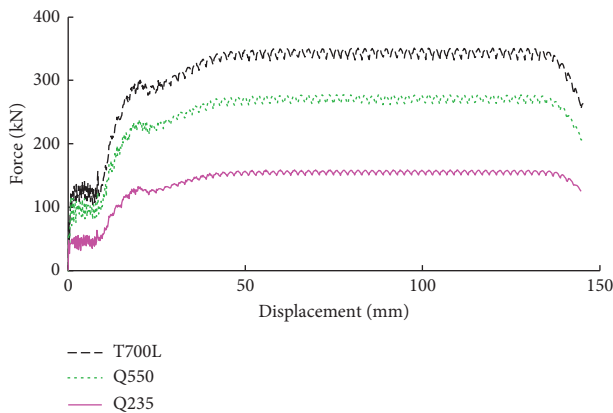


FIGURE 6: Material force-displacement.

through plastic deformation; second, the resulting conditions effectively improve the impact resistance of the anchor; third, the deformation space of the device provides a certain amount of energy release space for coal rocks. As a result, the surrounding rock impact energy is indirectly dissipated, the release and conversion of impact energy are effectively guided and controlled, and the impact energy is diminished in the buffering process of anchor bolts, thus

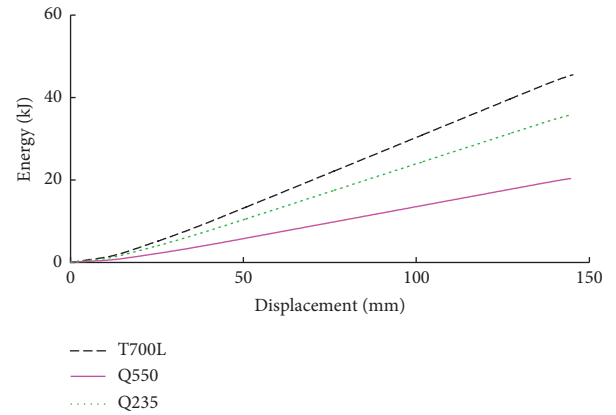


FIGURE 7: Energy absorption characteristics of materials.

ensuring the stability of the surrounding rock and support system of the roadway [29–31].

#### 4. Yielding Test Analysis of Energy-Absorbing Anti-Shock Anchor

The energy-absorbing device theoretically analyzed is made of Q235 material with an inner diameter, wall thickness, and length of 44, 3.5, and 150 mm, respectively. The size of the special-shaped nut is shown in Figure 2. The rod body is made of a 22 mm diameter rebar. Ordinary and energy-absorbing anchor bolts are subjected to static load and impact tests. These tests are performed on a 600 kN anchor bolt static load test bench (Figure 8) and a 300 kN anchor bolt impact test bench (Figure 9). The comparison of the energy-absorbing anchor before and after deformation is shown in Figure 10. The force-displacement curves of the common and energy-absorbing bolts tested are shown in Figure 11; Figure 12 depicts the energy absorption characteristic curves.

The tensile strength, yield distance, and absorbed energies of ordinary and energy-absorbing bolt are summarized in Tables 3 and 4.

- (1) Under static load, the tensile strength, yield distance, and absorption energy of ordinary anchor bolts are 505.46 MPa, 197 mm, and 31.99 kJ respectively, and those of the energy-absorbing anti-scour anchor bolts are 506.46 MPa, 329 mm, and 51.62 kJ, respectively. The tensile strength, yield distance, and absorption energy of the latter are 1, 1.67, and 1.61 times those of the former.

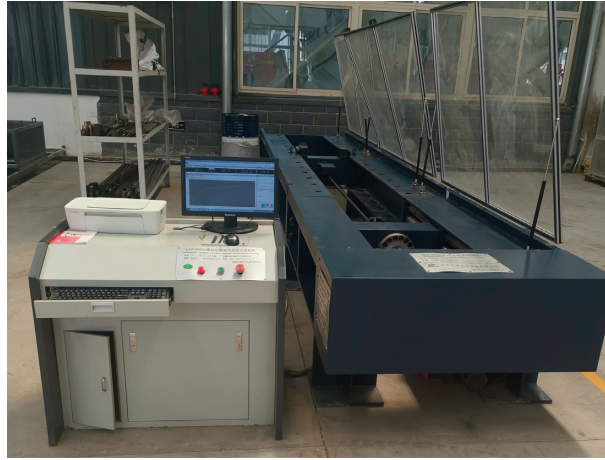


FIGURE 8: 600 kN test bench.



FIGURE 9: 300 kN test bench.

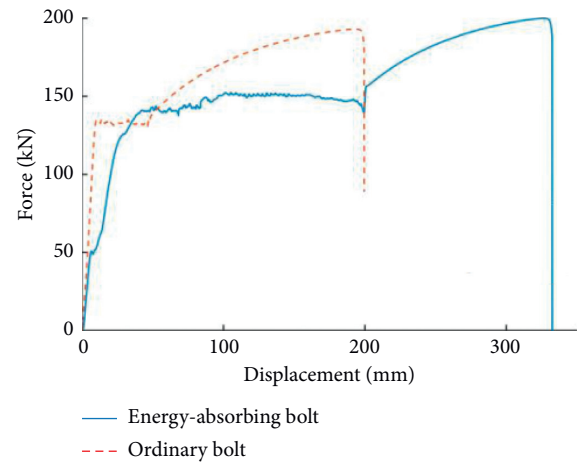


FIGURE 11: Force-displacement curves.



FIGURE 10: Before and after experiment contrast.

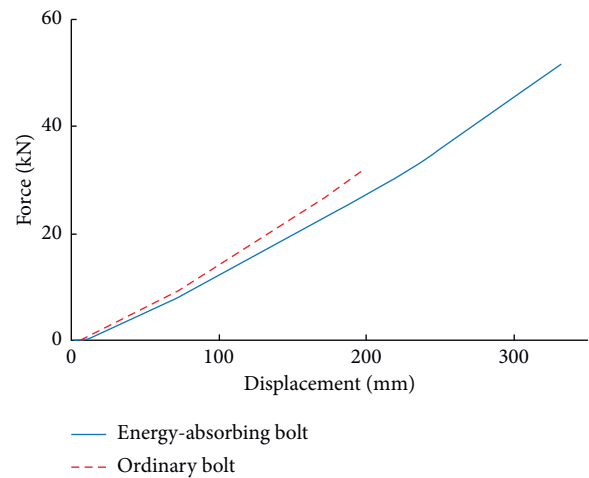


FIGURE 12: Energy absorption characteristics.

(2) Under impact load, the tensile strength, yield distance, absorbed energy, and impact time of ordinary bolt are 535.31 MPa, 195 mm, 35.08 kJ, and 0.15 s, and those of the energy-absorbing anti-shock bolt

are 534.84 MPa, 332 mm, 57.63 kJ, and 0.22 s, respectively. The tensile strength, yield distance, absorbed energy, and impact time of the latter are 1, 1.70, 1.64, and 1.47 times those of the former.

TABLE 3: Static loading of ordinary and energy-absorbing anti-shock bolt.

| Bolt form            | Tensile strength (MPa) | Abdicated distance (mm) | Absorbed energy (kJ) |
|----------------------|------------------------|-------------------------|----------------------|
| 335/22 ordinary bolt | 505.46                 | 197                     | 31.99                |
| MSGHS-335/22(1800)   | 506.55                 | 329                     | 51.62                |

TABLE 4: Impact loading of ordinary and energy-absorbing anti-shock bolt.

| Bolt form            | Tensile strength (MPa) | Abdicated distance (mm) | Absorbed energy (kJ) | Impact time (s) |
|----------------------|------------------------|-------------------------|----------------------|-----------------|
| 335/22 ordinary bolt | 535.31                 | 195                     | 35.08                | 0.15            |
| MSGHS-335/22 (1800)  | 534.84                 | 332                     | 57.63                | 0.22            |

- (3) Under the action of impact load, the tensile strength of both ordinary and energy-absorbing anti-impact bolts increase by 5%. The absorbed energy also increases, but the effect of loading mode on bolt displacement is reduced. This demonstrates that the latter has a stronger capacity to absorb energy and longer impact resistance time than the former.

## 5. Conclusion

The energy balance equation of the energy-absorbing support and roadway surrounding rock system is established based on the principle of energy conservation. Moreover, the energy criterion for the instability of roadway supported by an energy-absorbing system is deduced. It can be concluded that this type of support can significantly enhance the anti-scour performance of the support system and effectively prevent the occurrence of rockburst. Accordingly, the foregoing provides a theoretical basis for energy-absorbing support and anti-scour design.

In view of the advantages afforded by energy-absorbing supports, a type of energy-absorbing anti-scour bolt, composed of rod body, tray, constant resistance energy-absorbing device, and special nut, is designed and developed. Based on the theoretical and numerical analyses of mechanical properties, it is concluded that the energy-absorbing device has a repeatable deformation failure mode and a constant yielding force. The test results indicate that the energy-absorbing anti-scour bolt has a stronger impact mechanical properties than the ordinary bolt.

This paper presents the principle involved in energy absorption and anti-scour. The energy-absorbing support effectively guides and controls the release and transformation of impact energy. It also consumes the impact energy in the process of bolt displacement and buffer to ensure the stability of the surrounding rock and support system.

## Data Availability

All data, models, and codes that support the findings of this study are available from the corresponding author upon reasonable request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

## Acknowledgments

This study was supported by the National Natural Science Foundation of China (51804152), Liaoning Revitalization Talents Program (XLYC1907168), Natural Science Foundation of Liaoning Province (2019-MS-163), and Discipline Innovation Team of Liaoning Technical University (LNTU20TD08).

## References

- [1] M. He, G. Zhibiao, "Mechanical property and engineering application of anchor bolt with constant resistance and large deformation," *Chinese Journal of Rock Mechanics and Engineering*, vol. 33, no. 7, pp. 1297–1308, 2014.
- [2] C. Li, "A new energy-absorbing bolt for rock support in high stress rock masses," *International Journal of Rock Mechanics and Mining Sciences*, vol. 47, no. 2, pp. 396–404, 2010.
- [3] W. Hu and H. E. Manchao, *Status and Development Trends of Deep Coal Resources and Development Geological Conditions*, Coal Industry Publishing House, Beijing, China, 2008.
- [4] A. Wu, S. Chen, Y. Wang, and X. Chen, "Failure mechanism and supporting measures for large deformation of soft rock roadway in baluba copper mine," *Archives of Mining Sciences*, vol. 63, no. 2, pp. 449–464, 2018.
- [5] D. Yin, S. Chen, B. Chen, and Z. Xia, "Simulation study on strength and failure characteristics of Coal-Rock composite sample with coal persistent joint," *Archives of Mining Sciences*, vol. 64, no. 3, pp. 609–623, 2019.
- [6] F. S. Wong, "Uncertainties in FE modeling of slope stability," *Computers & Structures*, vol. 19, no. 5-6, pp. 777–791, 1984.
- [7] B. Zhao, *Rockburst and Prevention*, Coal Industry Publishing House, Beijing, China, 1995.
- [8] P. Xie and Y. Wu, "Deformation and failure mechanism and support structure technologies for Goaf-Side entries in steep multiple seam mining disturbances," *Archives of Mining Sciences*, vol. 64, no. 3, pp. 561–574, 2019.
- [9] H. Shi, H. Zhang, L. Song, and Y. Wu, "Variation of strata pressure and axial bolt load at a coal mine face under the effect of a fault," *Archives of Mining Sciences*, vol. 64, no. 2, pp. 351–374, 2019.
- [10] A. Pytlik, "Tests on hydraulic props equipped with yield valves at dynamic load modelling a rock burst," *Archives of Mining Sciences*, vol. 63, no. 2, pp. 477–489, 2018.
- [11] Z. Guo and X. Yang, Y. Bai, F. Zhou, and E. Li, "A study of support strategies in deep soft rock: the horsehead crossing roadway in daqiang coal mine," *International Journal of Mining Science and Technology*, vol. 22, no. 5, pp. 665–667, 2012.
- [12] W. D. Ortlepp, P. K. Kaiser and D. R. McCreath, "The design of support for the containment of rock burst damage in tunnels-

- an engineering approach," *Rock Support in Mining and Underground Construction*, pp. 593–609, Balkema, Rotterdam, Netherlands, 1992.
- [13] A. J. Jager, "Two new support units for the control of rockburst damage," in *Proceedings of the International Symposium on Rock Support*, pp. 621–631, Sudbury, Canada, July 1992.
- [14] C. C. Li and P. I. Marklund, Edited by B. Nilsen, Ed., "Field tests of the cone bolts in the Boliden mines," in *Fjellsprengningsteknikk/Bergmekanikk/Geoteknikk*, L. Hamref. K. Rohde et al., Eds., 12 pages, Oslo:Norsk Jord og Fjellteknisk Forbund, Oslo, Norway, 2004.
- [15] R. Varden, R. Lachenicht, and J. Player, "Development and implementation of the Garford dynamic bolt at the kanowna belle mine," in *Proceedings of the 10th Underground Operators Conference*, pp. 395–404, Australian Centre for Geomechanics, Launceston, Australia, April 2007.
- [16] M. Charette and M. Plouffe, "Roofex-results of laboratory testing of a new concept of yieldable tendon," in *Deep Mining'07*, Y. Potvin, Ed., pp. 395–404, Australian Centre for Geomechanics, Perth, Australia, 2004.
- [17] W. Pytel, P. Mertuszka, and K. Szeptun, "Ductile twisted rockbolt for underground excavation in deep mine conditions," *Instytutu Gospodarki Surowcami Mineralnymi I Energią Polskiej Akademii Nauk*, vol. 103, pp. 17–28, 2018.
- [18] Y. Pan and Z. Yang, "Numerical study on energy absorption of aluminum-composite hybrid tubes under axial quasi-static and impact crushing," *Journal of Vibration and Shock*, vol. 29, no. 8, pp. 209–213, 2010.
- [19] Z. Li, J. Yu, and L. Guo, "Experimental investigations on the energy absorption behavior of aluminum tubes with inductive structures subjected to axial loading," *Engineering Mechanics*, vol. 29, no. 6, pp. 346–352, 2012.
- [20] H. Wu, Z. Liang, F. Yang et al., "Energy absorption characteristics of the expansion of thin-walled circular metal tubes by middle-speed and high-speed impact of a rigid cylinder," *Journal of Applied Mechanics*, vol. 33, no. 2, pp. 325–331, 2016.
- [21] J. Yang, Z. Tang, F. He et al., "Energy absorption and anti-impact properties of mine diameter-expanding energy absorption components," *Journal of Vibration and Shock*, vol. 34, no. 8, pp. 134–138, 2015.
- [22] C. Lian, W. Xu, Y. Wang et al., "Numerical simulation of entry performance supported by a new high strength and high pretension yieldable bolts," *Rock and Soil Mechanics*, vol. 31, no. 7, pp. 2329–2335, 2010.
- [23] Z. Chen, D. Chen, Y. Zhang, X. Cheng, M. Zhang, and C. Wu, "Deep learning for autonomous ship-oriented small ship detection," *Safety Science*, vol. 130, 10 pages, 2020.
- [24] S. R. Dubey, S. K. Singh, and R. K. Singh, "A multi-channel based illumination compensation mechanism for brightness invariant image retrieval," *Multimedia Tools and Applications*, vol. 74, no. 24, pp. 11223–11253, 2015.
- [25] L. Dong, W. Wu, Q. Guo, M. N. Satpute, T. Znati, and D. Z. Du, "Reliability-aware offloading and allocation in multilevel edge computing system," *IEEE Transactions on Reliability*, vol. 1, 2019.
- [26] W. Wei, Q. Xu, L. Wang et al., "GI/geom/1 queue based on communication model for mesh networks," *International Journal of Communication Systems*, vol. 27, no. 11, pp. 3013–3029, 2014.
- [27] W. Wei, J. Su, H. Song, H. Wang, and X. Fan, "CDMA-based anti-collision algorithm for epc global c1 gen2 systems," *Telecommunication Systems*, vol. 67, no. 3, pp. 1–9, 2018.
- [28] D. Jiang, G. Li, Y. Sun, J. Kong, and B. Tao, "Gesture recognition based on skeletonization algorithm and CNN with ASL database," *Multimedia Tools and Applications*, vol. 78, no. 21, pp. 29953–29970, 2019.
- [29] W. Zheng-yi, L. -M. Dou, and G. -F. Wang, "Failuer mechanism of anchored bolts supporting structure of circular roadway under dynamic load," *Chinese Journal of Geotechnical Engineering*, vol. 37, no. 10, pp. 1901–1909, 2015.
- [30] H. -p. Kang, Y. -z. Wu, J. He et al., "Rock bolting performance and field practice in deep roadway with rock burst," *Journal of China Coal Society*, vol. 40, no. 10, pp. 2225–2233, 2015.
- [31] L. I. Chen, H. E. Man-chao, and W. -I. Gong, "Analysis on impact dynamics of negative poisson's ratio effect of anchor bolt with constant resistance and large deformation," *Journal of China Coal Society*, vol. 41, no. 6, pp. 1393–1399, 2016.

## Research Article

# Electroencephalogram of Happy Emotional Cognition Based on Complex System of Music and Image Visual and Auditory

Lin Gan,<sup>1</sup> Mu Zhang ,<sup>2</sup> Jiajia Jiang,<sup>1</sup> and Fajie Duan<sup>1</sup>

<sup>1</sup>School of Precision Instrument and Opto-Electronics Engineering, Tianjin University, Tianjin 300072, China

<sup>2</sup>College of Intelligence and Computing, Tianjin University, Tianjin 300350, China

Correspondence should be addressed to Mu Zhang; [sya@tju.edu.cn](mailto:sya@tju.edu.cn)

Received 23 October 2020; Revised 28 November 2020; Accepted 21 December 2020; Published 30 December 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Lin Gan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

People are ingesting various information from different sense organs all the time to complete different cognitive tasks. The brain integrates and regulates this information. The two significant sensory channels for receiving external information are sight and hearing that have received extensive attention. This paper mainly studies the effect of music and visual-auditory stimulation on electroencephalogram (EEG) of happy emotion recognition based on a complex system. In the experiment, the presentation was used to prepare the experimental stimulation program, and the cognitive neuroscience experimental paradigm of EEG evoked by happy emotion pictures was established. Using 93 videos as natural stimuli, fMRI data were collected. Finally, the collected EEG signals were removed with the eye artifact and baseline drift, and the *t*-test was used to analyze the significant differences of different lead EEG data. Experimental data shows that, by adjusting the parameters of the convolutional neural network, the highest accuracy of the two-classification algorithm can reach 98.8%, and the average accuracy can reach 83.45%. The results show that the brain source under the combined visual and auditory stimulus is not a simple superposition of the brain source of the single visual and auditory stimulation, but a new interactive source is generated.

## 1. Introduction

The study of audiovisual synchrony induced electroencephalogram (EEG) is an important part of the brain-computer interface (BCI) system. The BCI system provides a new way for humans to communicate without relying on peripheral nerves and muscles, that is, to communicate and control the outside world by measuring brain waves or other electrophysiological signals in the human brain. By studying the cognitive mechanism of the brain, exploring its efficient way of information processing, and popularizing the research results in the fields of intelligent systems such as computers, the intelligent processing ability of computers can be improved, so as to promote the rapid development of information science.

By converting the original brainwave signal into characteristic space in the form of string to a certain extent, the signal noise can be reduced, the change mode of brainwave signal can be abstracted, and the local information of

brainwave signal can be kept to the maximum extent. In order to overcome the shortcomings of traditional methods, we extract the modal functions related to specific EEG signal tasks, which greatly improves the performance of emotion classification tasks based on EEG signals.

In the practical application of EEG signal analysis and processing, signal analysis methods such as time domain and frequency domain are successively introduced. Lawhern believes that the BCI uses neural activity as a control signal to communicate directly with the computer. For a given BCI example, the feature extractor and classifier are customized for the different features of the EEG control signal expected to limit its application to this specific signal. Here, he proposed whether it is possible to design a CNN architecture to accurately classify EEG signals from different BCI paradigms while making it as compact as possible. In this work, he introduced EEGNet, a compact convolutional network based on EEG's BCI. He introduced the use of deep convolution and separable convolution to construct an EEG-

specific model, which encapsulates the famous EEG feature extraction concept of BCI. He compared EEGNet with the current state-of-the-art methods of the four BCI paradigms: P300 visual evoked potential, error-related negative response (ERN), motor-related cortical potential (MRCP), and sensorimotor rhythm (SMR). Although the conclusions of his research are correct, the research objects are rather vague [1]. Zhang believes that regularization has become a way to prevent overfitting of the brain-computer interface EEG classification. The effectiveness of regularization is usually highly dependent on the choice of regularization parameters usually determined by cross-validation (CV). However, CV imposes two main restrictions on BCI: (1) the user needs a large amount of training data; (2) it takes a relatively long time to calibrate the classifier. These restrictions will greatly reduce the practicality of the system and may cause users to be reluctant to use BCI. They introduced the sparse Bayesian method to classify EEG by using the Laplacian prior, SLaplace. Under the framework of Bayesian evidence, they use Laplacian priors to learn sparse discriminant vectors in a hierarchical manner. All required model parameters can be automatically estimated from the training data without the need for CV. Although their research is more targeted, it is not comprehensive enough [2]. Van believes that many variables in social sciences, physical sciences, and biological sciences (including neuroscience) are nonnormally distributed. In order to improve the statistical properties of such data or to allow parameter testing, logarithmic or logarithmic transformations are usually used. Box-Cox transformation or ad hoc method is sometimes used for parameters, and for these parameters, there is no known transformation that approximates normality. However, these methods are not always consistent with Gaussian. They discussed a transformation that maps the probability distribution to the normal distribution as much as possible and has precise consistency for the continuous distribution. To illustrate this point, the transformation was applied to the theoretical distribution and applied to quantitative electroencephalogram (qEEG) measurements from repeated recordings of 32 highly abnormal subjects. Consistency with Gauss is better than using logarithm, logarithm, or Box-Cox transformation. Their research lacks experimental data [3].

The main contributions of this paper are as follows: (1) an effective experimental paradigm of cognitive neuroscience is established, and the cognitive law of visual emotion is obtained, which improves the classification recognition rate; (2) the analysis and processing process of EEG signal is improved to promote the better integration of computer science and medicine; (3) the feature extraction and classification algorithm improvement of EEG data are carried out by using ERP technology.

## 2. Emotional Cognition and EEG Signals

*2.1. Emotional Cognition Based on Complex System.* Emotion is a kind of psychological evaluation made by the body for the things in the surrounding environment relative to its own needs. It can be seen that the cognitive process has a very important meaning for the generation and adjustment

of emotions. The body constantly uses cognitive mechanisms to evaluate and judge whether things in the surrounding environment can meet the body's adaptation needs, and on this basis, there are positive or negative emotional reactions. When the body adopts a more concise and clear cognitive structure and strategies, the way the body evaluates things will be relatively simple, and the emotional experience generated at this time is more likely to be in a strong state; and when the human body adopts more complex and changeable cognitive structures and cognitive strategies, the body will evaluate the surrounding things from multiple aspects and multiple levels, and the emotional experience generated at this time is more likely to tend to a mild state; that is, the complexity of the cognitive structure and the difference in cognitive strategies can greatly affect the generation and experience of emotions [4].

Complex system theory is a frontier direction in system science, and its main purpose is to reveal some dynamic behaviors that are difficult to explain with existing scientific methods. Different from the traditional reductionist method, the complex system theory emphasizes the combination of holism and reductionism to analyze the system. Complex systems are very sensitive to changes in individual parameters and partial structures in the system, while the human brain and nervous system are nonlinear and extremely complex systems. Therefore, when conducting research on emotional cognition, people will use the theory of complex systems to better feel the changes in the human body and achieve a high degree of correlation in many fields such as psychology, physiology, and neuroscience.

*2.2. EEG Signal.* When the external stimulation acts on the neurons, the potential difference between the inner and outer sides of the cell membrane decreases, and the excitability is enhanced. As the active potential is generated, a peak pulse will be generated on both sides of the cell membrane, which changes the positive and negative values from the inside to the outside. Synapses are the processes needed to transmit excitability between different neurons, and synapses play a very important role in transmitting excitability. In the process of transmission is excited, excited feeling of cells before contact position after pulse synapses, capsule into the hot spring active, after the release of the neurotransmitter substance synaptic relaxation, like a receiver cell membrane, it can be felt on the surface of the neurotransmitter substance and catch the body cause a series of changes of cell membrane ion channels, to cause the membrane potential of, namely the trigger synapses potential [5].

Brain waves can be roughly regarded as the dominant waveform of sine waves, so the waveform of brain waves can be represented by parameters such as frequency, amplitude, and phase. (1) Alpha waves appear when you are awake and close your eyes quietly. The frequency is 8~13 Hz, and the amplitude is 20~100  $\mu v$ . This is the most rhythmic waveform in brain waves. During visual stimulation or related cognition, the alpha wave will be immediately replaced by the beta wave. (2) Beta waves appear when the brain is excited,



which is related to mental tension and emotional excitement. The frequency is 14~30 Hz and the amplitude is 5~20  $\mu\text{V}$ , which is a fast wave. (3) When Cida waves indicate sleepiness or mental anxiety, the frequency is 4~7 Hz and the amplitude is 10~50  $\mu\text{V}$  when satisfied. (4) Delta waves are under deep anesthesia, hypoxia, or organic brain diseases, with a frequency of 0.5 to 3 Hz and an amplitude of 20 to 200  $\mu\text{V}$ . The actual measured brain wave is a signal composed of the above-mentioned multiple frequency components, which usually contains a lot of background noise [6].

The characteristics of EEG are as follows: (1) EEG is weak and interference noise is strong. Generally, the amplitude of EEG is only about 50  $\mu\text{V}$ . In the observation, the signal of the nonstudy object is very strong; for example, some unavoidable interference factors will cause strong interference noise. Due to these interference factors, the requirements for EEG extraction and processing devices need to be increased. For example, EEG detection systems and analysis systems require high input impedance, high common mode removal ratio, and low noise amplification technology. (2) EEG is unstable and random. The instability of the signal means that the statistical characteristics of the signal have nothing to do with the time of statistical analysis. In fact, the rhythm of brain waves is related to changes in mental state. The nonstationarity of brain waves is caused by certain changes in the physiological factors that constitute brain waves, and it has a relatively strong ability to automatically adapt to the outside world. (3) The frequency domain characteristics of EEG are clear, and power spectrum analysis and various frequency processing technologies decide to occupy a more important position than other physiological telecommunications. (4) There is very important mutual information between each read signal. This is because EEG generally uses multichannel signals obtained by a multielectrode measuring device [7].

**2.3. Emotional Classification of EEG Signals.** Emotional brainwave signals are generated by test subjects under specific emotional stimuli. Compared with sleep-related brain wave analysis tasks, the brain wave signal generated by emotional stimulation is longer. All points of the entire brain wave signal have nothing to do with a specific emotion, and most of the emotions are generated in local parts [8].

The Gaussian process is a collection of probability variables, which are distributed according to the combined Gaussian. In the Gaussian process regression, these probability variables represent the values of independent variable functions. Gaussian process regression assumes that the average value of the independent variable function distribution is 0, and the correlation between them is represented by the covariance function. The commonly used covariance function is as follows [9]:

$$K_{i,j} = K(X_i, X_j) = \theta_0 \exp \left\{ -\frac{1}{2} \sum_{d=1}^{D_1} \eta_d (x_{i,d} - x_{j,d}) \right\}. \quad (1)$$

The fully connected layer usually uses the Softmax model to solve multiclassification problems. The loss function of Softmax is as follows:

$$J(\theta) = -\frac{1}{m} \left[ \sum_{i=1}^m \sum_{j=1}^k l\{y^{(i)} = j\} \log \frac{e^{\theta_j^i}}{\sum_k e^{\theta_k^i}} \right]. \quad (2)$$

In the formula,  $\theta_j^i$  represents the input of the  $j$ th neuron node of the  $l$ th layer (usually the last layer), and  $\sum_k e^{\theta_k^i}$  represents the sum of the inputs of all the neuron nodes of the entire  $l$  layer. In order to prevent the local optimization of  $J(\theta)$ , the weight attenuation term is introduced. The specific expression is as follows:

$$J(\theta) = -\frac{1}{m} \left[ \sum_{i=1}^m \sum_{j=1}^k l\{y^{(i)} = j\} \log \frac{e^{\theta_j^i}}{\sum_k e^{\theta_k^i}} \right] + \frac{\lambda}{2} \sum_{i=1}^m \sum_{j=0}^n \theta_{i,j}^2. \quad (3)$$

Human beings do not recognize things by pixel by pixel but get local information from a part of the area, collect all the local information, and finally integrate it into the global information. In general, in any image, the closer the distance between two pixels is, the greater the correlation between them will be, while the correlation between two pixels that are far away is relatively small [10]. In fact, human neurons can only capture the local information of a picture and do not respond to global information. But in the end, the currently selected attribute is classified according to the impurity function. If selecting this attribute can reduce the impurity, then this attribute can separate the data. If the impurity function is denoted as  $i(t)$ , the purity gain is denoted as

$$\Delta i(t) = i(t_p) - E[i(t_c)] = i(t_p) - P_l[i(t_l)] + P_r[i(t_r)]. \quad (4)$$

Dice similarity coefficient calculation formula is as follows:

$$\text{DICE} = \frac{2|\text{SEG} \cap \text{GT}|}{|\text{SEG}| + |\text{GT}|}, \quad (5)$$

$$y = \alpha W y + \beta_1 X - W \beta_2 X + \varepsilon, \quad (6)$$

$$\ln g d p_{it} = a_0 + a_1 du^* dt + \sum_{i=1}^N b_j X u + \varepsilon_u. \quad (7)$$

Video coefficient calculation formula is as follows:

$$\text{SEN} = \frac{|\text{SEG} \cap \text{GT}|}{\text{GT}}, \quad (8)$$

$$l_{\text{ssim}} = 1 - \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (9)$$

$$\psi = \sum_{x=1}^{\theta} V x = \sum_{x=1}^{\theta} \left( \frac{W x}{\sum_1^n W \mathfrak{S}} S x \right). \quad (10)$$

VOE coefficient calculation formula is as follows:

$$\text{VOE} = 1 - \frac{|\text{SEG} \cap \text{GT}|}{|\text{SEG} \cup \text{GT}|}, \quad (11)$$

$$U_2 = \begin{cases} s - p_1 - kx_2, \\ x - p_2 - k(1 - x_2). \end{cases} \quad (12)$$

RVD coefficient calculation formula is as follows:

$$\text{RVD} = 100\% \times \left( 1 - \frac{|\text{SEG}| - |\text{GT}|}{|\text{GT}|} \right). \quad (13)$$

Jaccard coefficient calculation formula is as follows:

$$\text{Jaccard} = \frac{|\text{SEG} \cap \text{GT}|}{|\text{SEG} \cup \text{GT}|}, \quad (14)$$

$$x_H = \frac{p_2 - p_1 + 1}{2}. \quad (15)$$

**2.4. Physiological Basis of Music- and Image-Induced EEG.** There is a close relationship between emotion and emotion, which is both difference and connection. The process of cognition is accompanied by human emotions and emotions. It comes from the process of cognition and affects the conduct of cognition and activities.

When we have different emotions such as joy, anger, and sadness, we usually have emotions first and then emotions. The generation of emotion does not need conditioned reflex, while emotion is gradually acquired and evolved in the society. Emotions are extremely unstable and emotions are relatively stable. Generally, the emotion is not stable, which is situational and temporary. Compared with emotion, emotion is more stable, which is an embodiment of essence and will not change at any time. Inducing different emotions is the most important premise of emotion research. Emotion can be induced by external stimulation and internal response. At present, the common methods of emotion induction can be divided into two types: subjective induction and event induction. The subject elicitation is to make the subjects recall the memory fragments with emotional color or imagine the scene with a specific emotional state to induce the specific emotions of the subjects. The disadvantage of this method is that it cannot ensure that the participants can recall the corresponding memory fragments or imagine the scene of a specific emotional state, so it is difficult to ensure that the subjects can successfully induce specific emotions; even if the induction is successful, it is difficult to measure the duration of the corresponding emotions. Event elicitation is based on the mirror neuron theory, using external means to induce subjects to produce corresponding emotions. It is the most common event inducing method to induce the subjects' different emotions through external stimuli such as pictures, music, and video, and it is also the most commonly used emotion-inducing method by researchers. Emotion induction is the precondition of emotion recognition research. If the subjects' emotion cannot be

induced successfully, the follow-up research will not be conducted or the wrong results will be obtained [11, 12].

### 3. EEG Experiment on Emotional Cognition under Visual and Auditory Synergy Stimulation

**3.1. Experimental Data Set.** The DEAP data set includes a preprocessed version of the original EEG signal, and its main content is shown in Table 1. This version downsamples the original signal, and the sampling frequency is after 128 Hz. The signal is filtered at 4 to 45 Hz through a bandpass filter, and then the traces of the electrical signal are removed by blind source separation. In this experiment, 32 induced EEG signals in the data set were set to 5.0, and joy, arousal, and excitement were classified as low (score <5.0) and high (score ≥5.0), respectively; in 3 emotional dimensions, two classifications were made separately [13].

**3.2. Experimental Platform.** The experiment uses presentation to write the experimental stimulation program. Presentation can interact well with ERP, MEG, fMRI, and so on and is often used for stimulus presentation and experimental process control in cognitive experiments. It runs under the Windows environment and can reach millisecond time accuracy [14].

#### 3.3. Experimental Process

- (1) fMRI data collection was performed using 93 videos as natural stimuli. These videos were divided into eight large segments and played to three subjects using an MRI-compatible VR eyewear device, while they were scanned by fMRI. Parameters were 30-axis slice, matrix size 64 × 64, layer thickness 4 mm, 220 mmFOV, TR = 1.5 s, TE = 25 ms, and ASSET = 2 [15].
- (2) In order to prevent subjects from being stimulated by strong picture colors during the experiment, the experiment adopted a black background with a resolution of 640\*480 and a size equal to half of the screen. Moreover, the experimental pictures had the same brightness and contrast. The pictures were randomly divided into 5 sections and presented in a random manner to effectively avoid the practice effect and fatigue effect. Each image presented 3 s, and images were continuously presented between stimuli in each bar. After the end of each section of the experiment, the subjects could choose whether to take a rest. The experiment lasted for 5 minutes [16].
- (3) EEG data acquisition: firstly, set the storage path of EEG data; then the EEG signals recorded on the screen of the subjects were observed. When the EEG signals stabilized, the audiovisual stimulation experimental paradigm was presented on the screen of the subjects. Finally, in accordance with the designed experimental paradigm flow, the subjects were induced by emotional pictures and sounds to collect EEG signals [17].

TABLE 1: Experimental data set.

|                        |                            |
|------------------------|----------------------------|
| Number of participants | 32                         |
| Number of videos       | 40                         |
| Video duration         | 60 s                       |
| Emotional dimension    | Pleasure                   |
|                        | Arousal                    |
|                        | Like degree                |
|                        | Dominance                  |
| Value range            | Familiarity: 1.0 to 5.0    |
|                        | Others: 1.0 to 9.0         |
| Recorded signal        | Number of channels: 40     |
| Number of participants | Sampling frequency: 128 Hz |

### 3.4. Data Processing

- (1) Data format conversion (convert .cnt to .mat format): since the original data is a .cnt format file collected by the Scan software, this article uses the EEGlab toolbox to import it into Matlab and then process it. The data saved after processing is in the .mat format, and the data in the .mat format can be used. The data generated by subsequent feature extraction and other works can also use the .mat format.
- (2) Complete the independent component analysis processing of the data. Use the ICA algorithm in the EEGlab toolbox to decompose the data, and remove the artifact signals such as ocular electricity and myoelectricity to achieve denoising [18].

## 4. EEG Analysis of Emotional Cognition under Visual and Auditory Costimulation Based on Complex System

**4.1. Analysis of Global Feature Difference Results.** First, select the initial value of the weight vector in the convolutional layer, the learning rate is  $1e-5$ , and the momentum factor is 0.9. Under the eight probability distribution conditions of uniform distribution, zero-point distribution, and normal distribution, the effects of different initialization weight vectors on the accuracy of emotion recognition are shown in Figure 1. It can be found from the table that the classification accuracy is the highest when the initialization weight vector is uniformly distributed. The power spectrum entropy of  $P$  area,  $PT$  area, and  $O$  area does not change significantly with the level, which also shows that these areas have a small relationship with emotions.

As shown in Figure 2 and Table 2, the power spectrum entropy of  $F$  zone,  $AT$  zone, and  $C$  zone fluctuates greatly with the level change, especially in  $F$  zone. And the power spectrum entropy of the  $F$  zone has an upward trend with the level change. This may be due to the fact that the brain is in a highly stressed state when viewing lower-level pictures, which makes the regularity of the brain waves stronger, so that the power spectrum entropy is higher.

As shown in Table 3, as the level of the picture increases, the state of the subjects is more relaxed when watching the

picture, which makes the law of brain waves weaker, and thus the power spectrum entropy becomes larger. The importance of emotion to human life is self-evident, and it affects the individual's evaluation of external things and affects the individual's behavior mode in dealing with external things. The relationship between emotion and executive function has also become a research hotspot in recent years. Through research on patients with anxiety and depression, people have found that emotions affect individual working memory, renewal, transformation, and other abilities [19, 20].

Figure 3 shows the accuracy comparison of the two-category emotion recognition of the five algorithms. The horizontal axis in the figure represents the tester number, and the vertical axis represents the accuracy rate. It can be seen from the figure that when using statistical features as input, two algorithms, RBF-SVM and Linear-SVM, are used for emotion recognition, and their accuracy is generally lower than that of the algorithm in this paper.

As shown in Table 4, through the comparison of the accuracy of the algorithm in this paper and the algorithm in the literature, it can be found that this paper uses the preprocessed brain wave signal as input. By adjusting the parameters of the convolutional neural network, the highest accuracy rate of the two-class algorithm can reach 98.8%, and the average accuracy rate is 83.45%; compared to the emotion recognition algorithm of the convolutional neural network that uses statistical features as input, it is significantly improved [21].

As shown in Figure 4, with the increase in sparsity, the clustering coefficient  $C$  under different sparsity levels increases monotonically, and the clustering coefficient can measure the clustering characteristics and tightness of the brain function network. Under high arousal conditions, the clustering coefficient of the brain network with low pleasure is smaller than the brain network of high pleasure; under the condition of low pleasure, the clustering coefficient of low arousal is greater than the clustering coefficient of high arousal.

As shown in Figure 5 and Table 5, this may indicate that, under higher pleasure conditions, the brain nerves are more excited, local connections between brain regions increase, and brain function connections are enhanced, and under the same conditions of low pleasure, the brain function of the brain with low arousal is higher than the high arousal. The brain function of the brain network has more connections between brain regions, and the local connection function is stronger. The research results on local efficiency are consistent with the research results of clustering coefficients. Both of these two attributes can reflect the differentiation ability of local brain functions. Together, the two indicate that high-pleasure emotions promote the local information processing and processing of the brain [22].

**4.2. Analysis of EEG Feature Extraction Results.** The average accuracy of discrete dimensions is shown in Table 6. As can be seen from the figure, the average accuracy has improved. When using the convolutional neural network to solve the

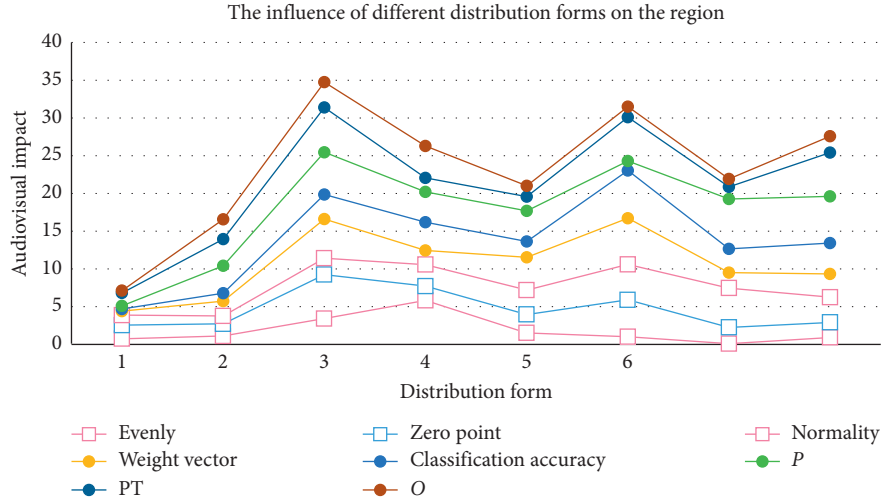


FIGURE 1: The influence of different distribution forms on the region.

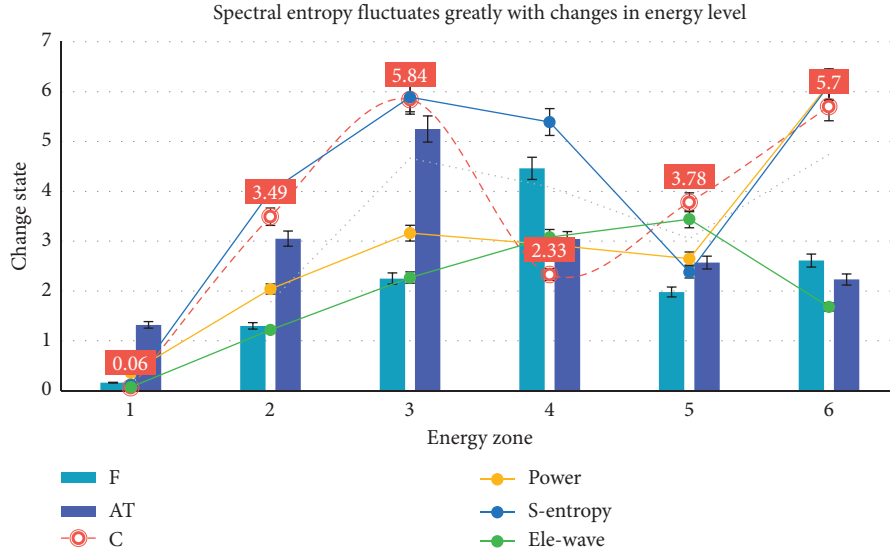


FIGURE 2: Spectral entropy fluctuates greatly with changes in energy level.

TABLE 2: The influence of different energy fluctuation forms on the accuracy of emotion recognition.

| Number | $F$  | AT   | C    | Power | S-entropy | Ele-wave |
|--------|------|------|------|-------|-----------|----------|
| 1      | 0.16 | 1.32 | 0.06 | 0.36  | 0.12      | 0.07     |
| 2      | 1.3  | 3.05 | 3.49 | 2.04  | 4         | 1.22     |
| 3      | 2.25 | 5.25 | 5.84 | 3.16  | 5.89      | 2.27     |
| 4      | 4.46 | 3.04 | 2.33 | 2.93  | 5.39      | 3.08     |
| 5      | 1.98 | 2.57 | 3.78 | 2.65  | 2.38      | 3.44     |
| 6      | 2.61 | 2.23 | 5.7  | 6.15  | 6.13      | 1.68     |

problem of image classification, the input is unstructured image data; that is, we do not need to extract all the features of an image. At the same time, the operation of each layer in the convolutional neural network is equivalent to extracting the image features with a feature operator, and the parameters of the feature operator can be adjusted and updated

TABLE 3: The influence of different initialization weight vectors on the accuracy of emotion recognition.

| Data distribution | Accuracy | Variance |
|-------------------|----------|----------|
| Uniform           | 0.636827 | 0.013589 |
| Lecun_uniform     | 0.601238 | 0.010331 |
| Normal            | 0.619266 | 0.094318 |
| Zero              | 0.558342 | 0.024774 |
| Glorot_normal     | 0.571085 | 0.043303 |
| Glorot_uniform    | 0.571985 | 0.025202 |
| He_normal         | 0.593892 | 0.012603 |
| He_uniform        | 0.590875 | 0.051937 |

continuously during the whole network training process, so as to optimize the classification results.

As shown in Figure 6, the number of features that can be extracted in the whole network is closely related to the scale

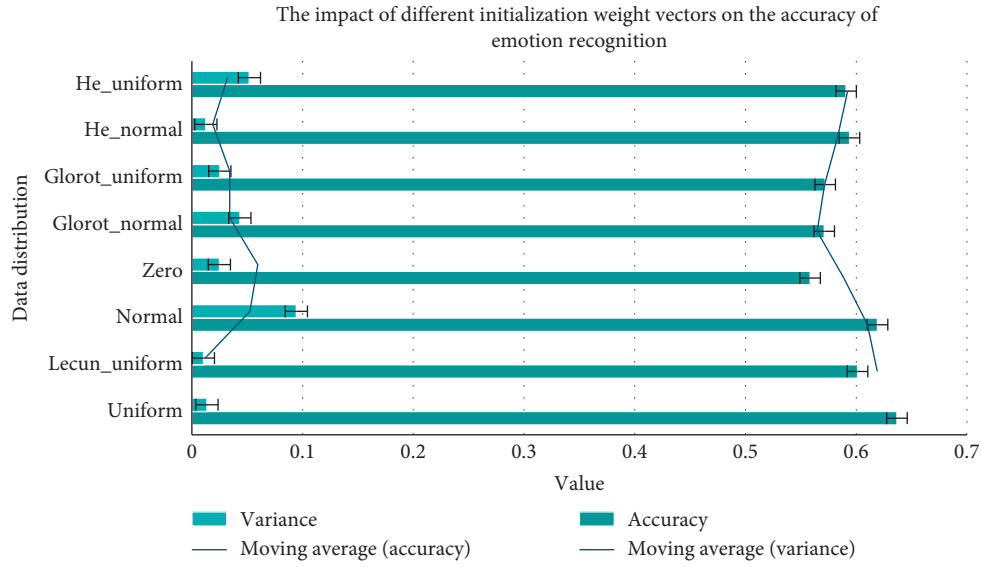


FIGURE 3: The impact of different initialization weight vectors on the accuracy of emotion recognition.

TABLE 4: The parameters of the convolutional neural network.

| Item        | Stimulate | Neurons | Cell membrane | Polarization | Elimination state |
|-------------|-----------|---------|---------------|--------------|-------------------|
| Positive    | 1.63      | 1.76    | 0.8           | 1.99         | 1.02              |
| Overlapping | 2.65      | 1.94    | 3.32          | 3.68         | 1.28              |
| Galvanic    | 4.15      | 2.18    | 4.58          | 3.7          | 2.83              |
| pF          | 2.33      | 3.53    | 4.26          | 1.77         | 5.86              |
| LO          | 3.42      | 4.44    | 3.61          | 2.12         | 3.86              |
| PPA         | 6.32      | 2.84    | 3.09          | 5.63         | 4.65              |

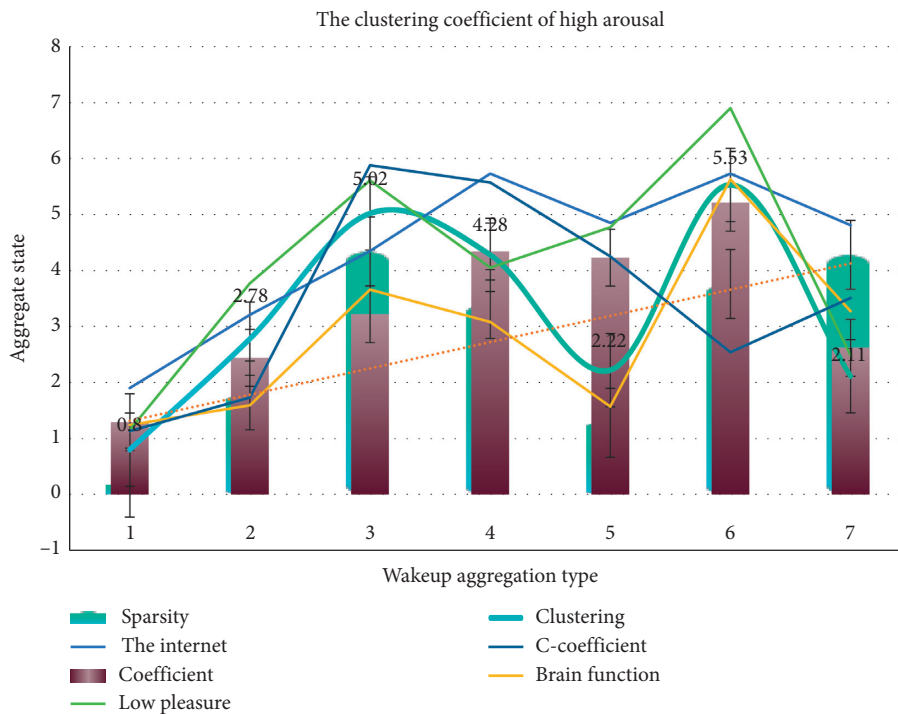


FIGURE 4: the clustering coefficient of high arousal.

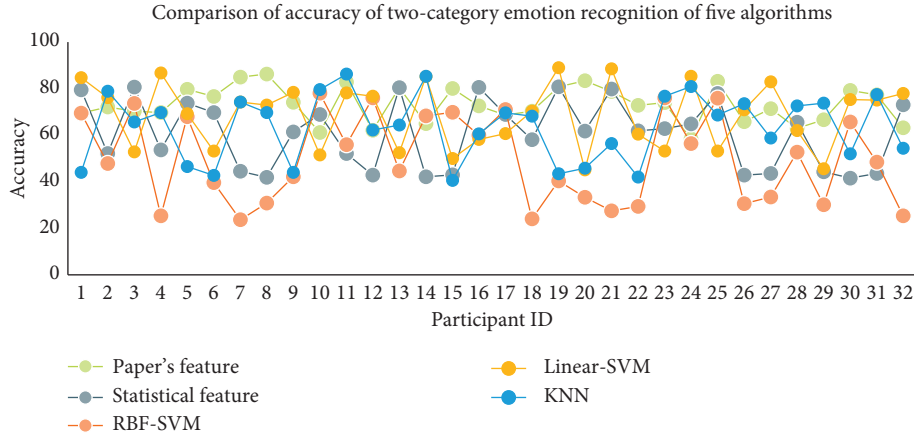


FIGURE 5: Comparison of accuracy of two-category emotion recognition of five algorithms.

TABLE 5: The influence of different emotional states on brain nerves.

| N | Sparsity | Clustering | Coefficient | Brain function | Internet | Pleasure | C-coefficient |
|---|----------|------------|-------------|----------------|----------|----------|---------------|
| 1 | 0.21     | 0.8        | 1.29        | 1.24           | 1.9      | 1.15     | 1.13          |
| 2 | 1.77     | 2.78       | 2.44        | 1.59           | 3.21     | 3.77     | 1.73          |
| 3 | 4.34     | 5.02       | 3.22        | 3.66           | 4.35     | 5.61     | 5.88          |
| 4 | 3.4      | 4.28       | 4.34        | 3.08           | 5.73     | 4.05     | 5.57          |
| 5 | 1.28     | 2.22       | 4.23        | 1.57           | 4.85     | 4.77     | 4.25          |
| 6 | 3.76     | 5.53       | 5.21        | 5.63           | 5.73     | 6.9      | 2.54          |
| 7 | 4.28     | 2.11       | 2.62        | 3.27           | 4.81     | 2.48     | 3.51          |

TABLE 6: Average accuracy of discrete dimensions.

| $F$   | $B$   | $I$   | $F+B$ | $F+I$ | $B+I$ | $F+B+I$ | $D$   |
|-------|-------|-------|-------|-------|-------|---------|-------|
| 28.72 | 24.36 | 20.18 | 17.67 | 24.86 | 16.20 | 21.70   | 21.04 |
| 26.88 | 40.58 | 29.49 | 37.71 | 46.18 | 38.23 | 52.15   | 31.71 |
| 30.38 | 30.90 | 45.35 | 35.69 | 28.86 | 57.28 | 26.90   | 45.66 |
| 40.26 | 43.93 | 53.85 | 32.44 | 47.60 | 30.04 | 30.68   | 45.14 |
| 47.23 | 56.52 | 57.87 | 43.79 | 46.10 | 33.78 | 44.57   | 45.12 |
| 28.54 | 48.68 | 26.68 | 32.22 | 49.40 | 47.79 | 27.04   | 43.00 |
| 27.40 | 54.36 | 59.09 | 34.21 | 26.55 | 27.69 | 58.40   | 35.42 |
| 38.61 | 58.39 | 57.08 | 45.20 | 59.02 | 43.64 | 41.72   | 60.09 |
| 47.44 | 43.05 | 26.50 | 48.31 | 31.87 | 57.51 | 52.14   | 38.41 |
| 53.43 | 49.12 | 41.07 | 30.83 | 53.13 | 38.85 | 45.43   | 45.77 |

of the network model. The larger the scale of the network model is, the more features can be extracted and the more types can be effectively distinguished [23]. In addition, since the convolution kernel parameters are constantly updated and adjusted throughout the training process, there is no need to pay special attention to the processing results of each layer, but to adjust the parameters of each layer within a reasonable range based on the output error of the network, and finally complete. From autonomous learning to abstract feature expression, image classification is realized. Instead, we use the average intensity of all voxels in the selected brain area as a benchmark. Such a choice has an advantage over the same conversion factor in the expansion result.

As shown in Figure 7 and Table 7, compared to the gray matter intensity calculated separately, it affects the change in signal percentage. A brain with 35% white matter is 15% brighter than gray matter, and the average

signal percentage using the whole brain is reduced by 5% of the baseline using only gray matter. Generally, the magnitude of this difference is small compared to other errors. Similarly, the threshold used to cover the whole brain only has a small effect on the reasonable range of the threshold [24].

**4.3. Synchronous Analysis of EEG Induced by Visual and Auditory Perception.** The results of the EEG phase synchronization index  $t$ -test are shown in Figure 8 and Table 8. The results show that there are significant differences between the EEG synchronicity index of happy emotion cognition (positive emotion) and anger, sadness, surprise, disgust, and fear emotion cognition (negative emotion). When receiving pleasant facial expression pictures and sound synchronization stimulation, the parietooccipital lobe and the left frontal lobe have a high EEG phase synchronization; that is, the visual area and the left emotional area have a synchronized oscillation phenomenon, and the synchronization of the glume lobe and the left frontal lobe is not high; that is, the phenomenon of synchronous oscillation is not obvious.

As shown in Figure 9 and Table 9, it can be inferred that the visual channel plays a dominant role in the perception of pleasure. When receiving the stimulus of the sad expression picture and the sound synchronization, the occipital lobe and glume lobe have a high EEG phase synchronization with the right frontal lobe; that is, the visual area and auditory area have synchronized oscillations with the right emotional area, but the leaves are displayed. Compared with the

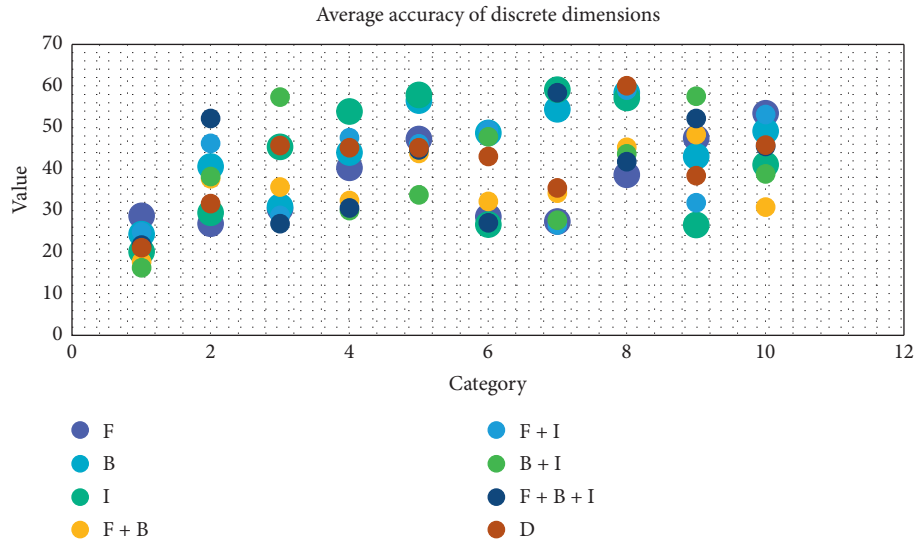


FIGURE 6: Average accuracy of discrete dimensions.

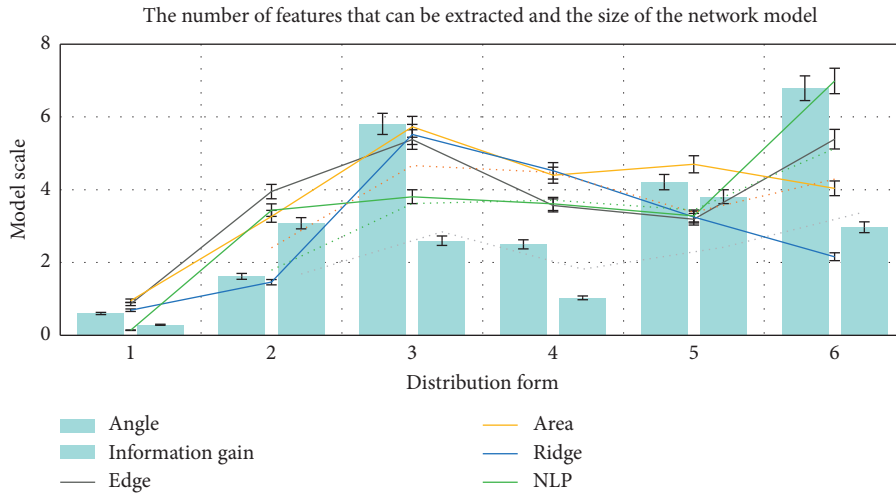


FIGURE 7: The number of features that can be extracted and the size of the network model.

TABLE 7: Comparison of the effects of gray matter intensity and white matter on the brain.

| $N$ | Angle | Edge | Information gain | Area | Ridge | NLP  |
|-----|-------|------|------------------|------|-------|------|
| 1   | 0.6   | 0.86 | 0.29             | 0.95 | 0.69  | 0.14 |
| 2   | 1.62  | 3.95 | 3.08             | 3.27 | 1.46  | 3.44 |
| 3   | 5.81  | 5.38 | 2.6              | 5.73 | 5.52  | 3.81 |
| 4   | 2.5   | 3.57 | 1.03             | 4.4  | 4.52  | 3.61 |
| 5   | 4.21  | 3.19 | 3.81             | 4.7  | 3.25  | 3.28 |
| 6   | 6.79  | 5.39 | 2.97             | 4.04 | 2.16  | 6.99 |

synchronization of the right frontal lobe, the synchronization between the occipital lobe and the right frontal lobe is higher. From this, it can be inferred that, in the cognition of sad emotions, the auditory channel plays a leading role and the visual channel plays a supporting role. When neurons are in a resting state, positive and negative charges are evenly

distributed inside and outside the cell membrane, but due to the overlap of the positive and negative centers, all neurons neither will exhibit electrical properties on the outside nor will form galvanic couples [25,26].

As shown in Figure 10 and Table 10, when a neuron is stimulated, the cell membrane is in a state of polarization cancellation, the internal and external charge distribution are uneven, and the centers of positive and negative charges do not overlap, thus forming a pair of galvanic couples. Under sound conditions, a similar situation occurs. The three brain regions pF, LO, and PPA form a network, but they have nothing to do with STS. STS can perceive three modal stimuli: single vision, single hearing, and audiovisual integration, so it is possible that this mode of sound is only processed in STS and has nothing to do with the network formed by those three brain regions. According to the results of multivoxel mode analysis, we see that the STS brain area is also involved in representing

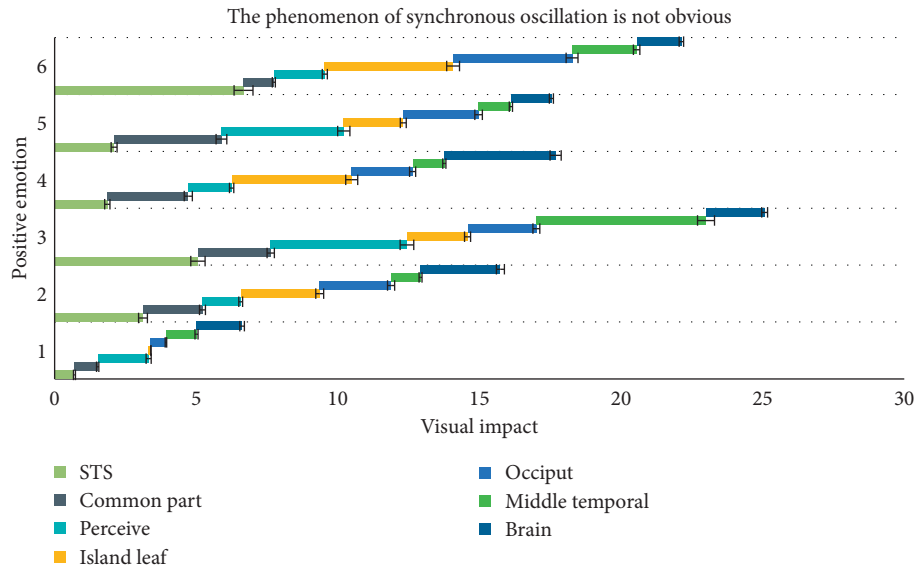


FIGURE 8: EEG synchronicity index of happy emotion cognition.

TABLE 8: Synchronized oscillation phenomenon.

| Visual channel | Pleasure | Sound synchronization | Glume leaf | Right frontal lobe | Auditory channel |
|----------------|----------|-----------------------|------------|--------------------|------------------|
| 1.14           | 0.46     | 1.46                  | 1.61       | 1.78               | 1.69             |
| 2.58           | 1.88     | 2.08                  | 2.08       | 3.57               | 2.1              |
| 5.79           | 4.93     | 5.46                  | 2.9        | 3.72               | 3.98             |
| 4.43           | 4.41     | 3.59                  | 4.97       | 4.61               | 2.26             |
| 3.87           | 1.15     | 1.19                  | 4.8        | 1.88               | 3.27             |
| 3.6            | 4.13     | 1.75                  | 6.87       | 3.03               | 3.77             |

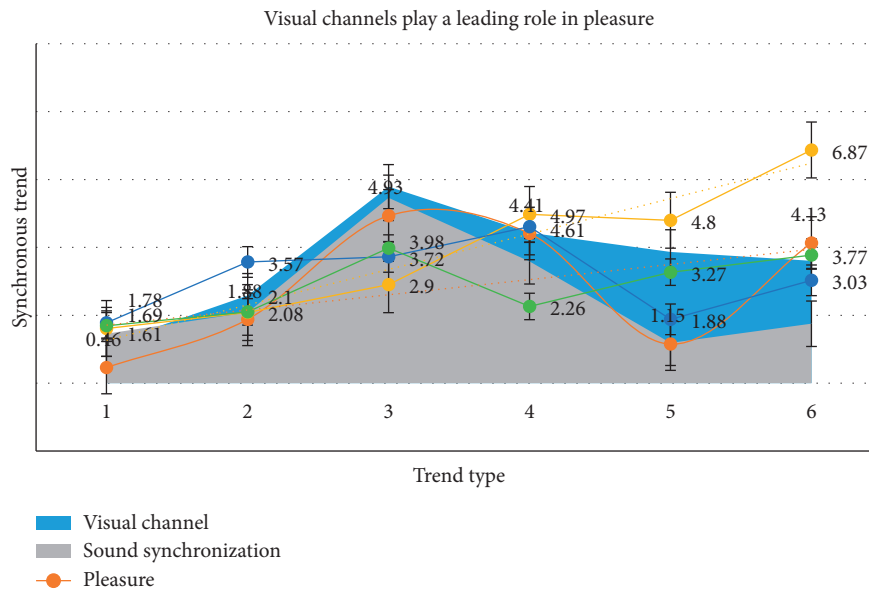


FIGURE 9: Visual channels play a leading role in pleasure.



TABLE 9: Visual channels play a leading role in pleasure.

| Item        | Stimulate | Neurons | Cell membrane | Polarization | Elimination state |
|-------------|-----------|---------|---------------|--------------|-------------------|
| Positive    | 1.63      | 1.76    | 0.8           | 1.99         | 1.02              |
| Overlapping | 2.65      | 1.94    | 3.32          | 3.68         | 1.28              |
| Galvanic    | 4.15      | 2.18    | 4.58          | 3.7          | 2.83              |
| pF          | 2.33      | 3.53    | 4.26          | 1.77         | 5.86              |
| LO          | 3.42      | 4.44    | 3.61          | 2.12         | 3.86              |
| PPA         | 6.32      | 2.84    | 3.09          | 5.63         | 4.65              |

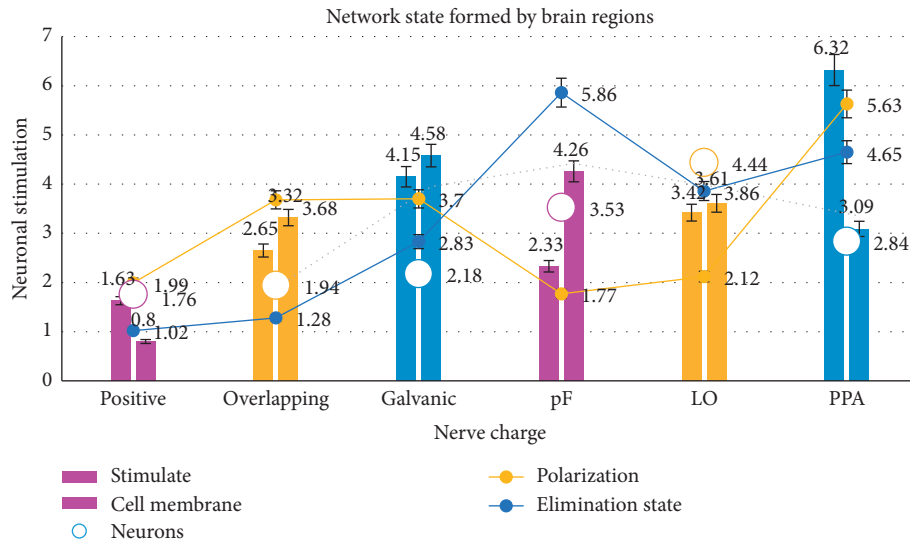


FIGURE 10: Network state formed by brain regions.

TABLE 10: The effect of charge neuron stimulation on vision.

| STS  | Common part | Perceive | Island leaf | Occiput | Middle temporal | Brain |
|------|-------------|----------|-------------|---------|-----------------|-------|
| 1.54 | 0.84        | 0.06     | 1.82        | 0.17    | 0.23            | 0.68  |
| 1    | 3.55        | 1.8      | 3.88        | 2.26    | 2.06            | 3.55  |
| 3.43 | 5           | 3.18     | 2.39        | 3.33    | 2.15            | 4.6   |
| 5.19 | 2.9         | 5.23     | 4.49        | 4.45    | 3.83            | 1.81  |
| 4.41 | 1.8         | 1.76     | 3.11        | 2.24    | 3.83            | 2.09  |
| 1.88 | 4.79        | 3.12     | 6.06        | 2.06    | 5.65            | 2.1   |

the semantic similarity between the scene and the sound, so we further explored the whole brain that is functionally connected to the STS under the sound and scene tasks.

As shown in Figure 11, the results found that there are common parts of the brain areas that are jointly responsible for processing scenes and sound tasks with STS: perception related (insular lobes), spatial imagery and memory (middle occipital region), semantic related (back of middle temporal gyrus) and object perception related brain area (LOC), and audiovisual integration. As the number of decomposition layers increases, the smoother the waveform is, the more the information is lost; on the contrary, the less the number of decomposition layers, the more the interference in the waveform. It can be obtained by plotting the approximation coefficients and detail coefficients of the above layers, and the preprocessed EEG signal is subjected to 5-layer wavelet decomposition, and the obtained waveform has good smoothness while retaining the details more completely.

Therefore, the wavelet coefficients of the fifth layer are algorithmically reconstructed to extract the P300 characteristics of evoked EEG in the visual, auditory, and visual and auditory target stimulation modes. Whether it is visual, auditory, or visual-auditory joint stimulation mode, the wavelet transform method can effectively extract the P300 characteristics of the evoked EEG signal, so that the target stimulus in the three modes can be effectively distinguished from the nontarget stimulus [27, 28].

As shown in Figure 12 and Table 11, in the ERP induced by different valence emotion pictures as a stimulus, the middle component and late component also showed significant differences in amplitude. In the right brain, the negative wave amplitude induced by high-pleasure pictures is significantly higher than the negative wave amplitude induced by low pleasure pictures. This phenomenon can indicate that positive stimulation can get more cognitive resources in the right brain, and emotions are in the right

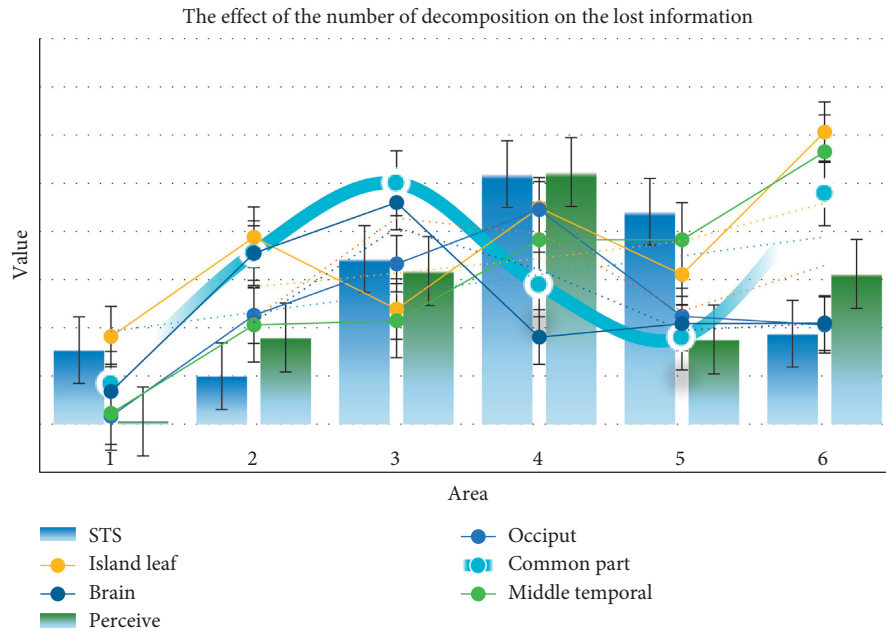


FIGURE 11: The effect of the number of decomposition layers on the lost information.

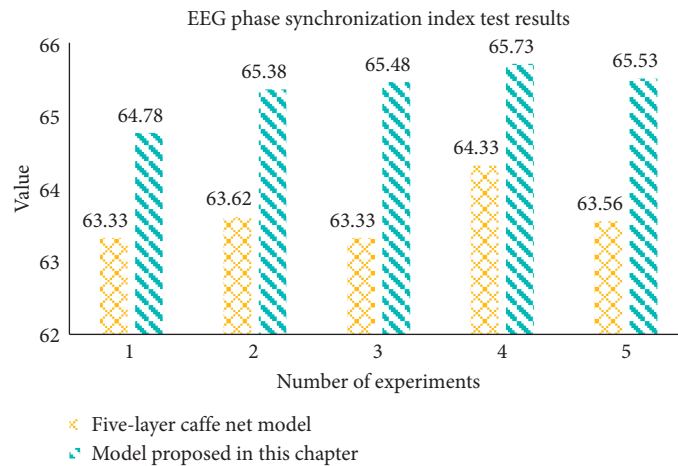


FIGURE 12: EEG phase synchronization index *t*-test results.

TABLE 11: The emotional effect of ERP fluctuations.

|             | Emotional processing | Emotional price | ERP role | Up negative | Negative stimulus |
|-------------|----------------------|-----------------|----------|-------------|-------------------|
| Left brain  | 1.42                 | 0.55            | 1.43     | 1.21        | 0.46              |
| Right brain | 1.44                 | 2.1             | 1.24     | 2.79        | 3.82              |
| Vi-info     | 5.7                  | 4.51            | 4.05     | 5.71        | 5.76              |
| Fron-lobe   | 4.28                 | 3.9             | 3.46     | 2.72        | 5.03              |
| Cent-area   | 1.7                  | 3.75            | 3.58     | 2.43        | 3.79              |

brain. Subject to more refined evaluation processing, this cognitive process involves attention and memory cognitive processes. Different processing of visual information with different valence exists in every stage of mental processing.

As shown in Figure 13, whether under high arousal conditions or low arousal conditions, after 300 ms, the left

and right brains have different processing methods for different valence of visual information, but they are all involved in the emotional processing process, both in the frontal and central regions. Reflect this difference in processing. The research on the ERP effect of emotional valence mainly focuses on negative bias, but there is no obvious

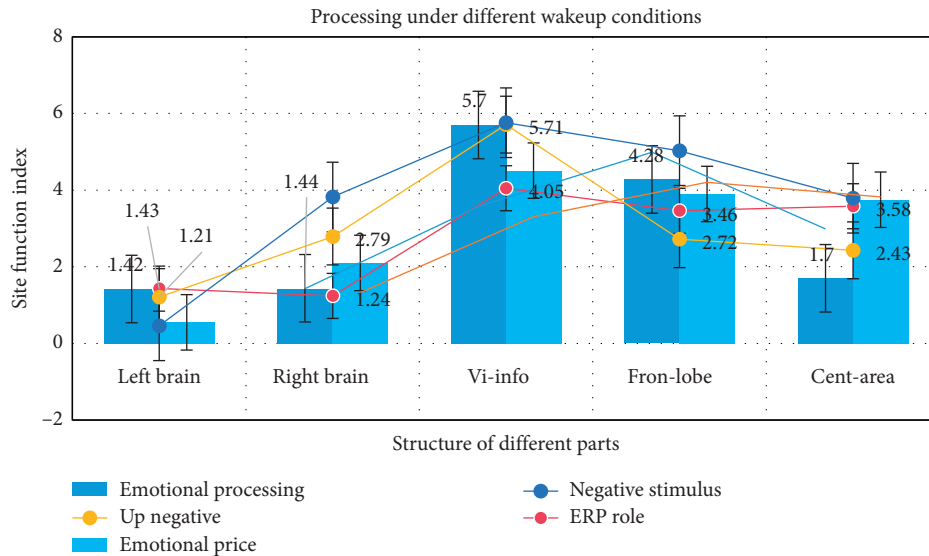


FIGURE 13: Processing under different wakeup conditions.

negative bias in this experiment. This is because negative bias mainly occurs when attention resources are relatively lacking, and negative stimuli are allocated to more psychological resources. When psychological resources are sufficient, there is no difference in advantage between positive and negative emotional information processing.

## 5. Conclusions

This paper mainly studies the changes of EEG under the synergy of music and image visual and auditory stimuli based on complex systems. The event-related potential technology is used to analyze and study the rules of visual emotion cognition. The brain activity analysis method based on emotional valence and the brain activity analysis method based on emotional arousal are used to compare the tester's emotions on the two indicators of valence and arousal. Marking as a label, by selecting music videos as the tester's emotional inducing material, it effectively induces the tester to produce obvious emotional changes. Let the tester watch the video intermittently, collect the physiological signals generated by the body during the viewing process, preprocess the collected human physiological signals to construct a new data set, and finally use the data set for the algorithm test.

Feature extraction is performed on the preprocessed visual, auditory, and visual-auditory joint evoked EEG signals. When a single channel (visual or auditory) stimulation is given to complete positive emotional cognition, the right frontal lobe has a higher synchronization with the visual or auditory areas of the brain, which is related to the brain receiving visual or auditory stimuli; while completing the negative during emotional cognition, the left frontal lobe and other functional areas did not show synchronized activities. The results of this study show that when receiving a single channel (visual or auditory) stimulation to complete emotional cognition, positive emotions can cause

synchronized electrical activity in related areas of the brain more than negative emotions.

The body's emotional state affects the pattern of attention distribution. If the body is in a positive emotional state, it is more likely to adopt a top-down approach to information processing, which relies on the body's already formed knowledge and experience structure. On the contrary, the organism in a negative emotional state is more likely to adopt a more systematic information processing strategy, which is a bottom-up information processing method, which does not depend on the knowledge and experience of the body. Meanwhile, the organism adopting this method will pay more attention to the details of the current stimulus.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] V. J. Lawhern, A. J. Solon, and N. R. Waytowich, "EEGNet: a compact convolutional network for EEG-based brain-computer interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, Article ID 056013, 2016.
- [2] Y. Zhang, G. Zhou, and J. Jin, "Sparse bayesian classification of EEG for brain-computer interface," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 11, 1 page, 2016.
- [3] S. J. Van Albada and P. A. Robinson, "Transformation of arbitrary distributions to the normal distribution with application to EEG test-retest reliability," *Journal of Neuroence Methods*, vol. 161, no. 2, pp. 205–211, 2018.
- [4] S. Olbrich, R. Van Dinteren, and M. Arns, "Personalized medicine: review and perspectives of promising baseline EEG

- biomarkers in major depressive disorder and attention deficit hyperactivity disorder,” *Neuropsychobiology*, vol. 72, no. 3-4, pp. 229–240, 2016.
- [5] Z. Lan, O. Sourina, L. Wang, and Y. Liu, “Real-time EEG-based emotion monitoring using stable features,” *The Visual Computer*, vol. 32, no. 3, pp. 347–358, 2016.
  - [6] G. D. Gennaro, P. P. Quarato, and F. Sebastiano, “Ictal heart rate increase precedes EEG discharge in drug-resistant mesial temporal lobe seizures,” *Clinical Neurophysiology*, vol. 115, no. 5, pp. 1169–1177, 2016.
  - [7] H. Merckelbach, P. Muris, and R. Horselenberg, “EEG correlates of a paper-and-pencil test measuring hemisphericity,” *Journal of Clinical Psychology*, vol. 53, no. 7, pp. 739–744, 2016.
  - [8] R. Chai, G. R. Naik, T. N. Nguyen et al., “Driver fatigue classification with independent component by entropy rate bound minimization analysis in an EEG-based system,” *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 3, pp. 715–724, 2017.
  - [9] W. O. Tatum, O. Selioutski, J. G. Ochoa et al., “American clinical neurophysiology society guideline 7: guidelines for EEG reporting,” *The Neurodiagnostic Journal*, vol. 56, no. 4, pp. 285–293, 2016.
  - [10] W. Biesmans, N. Das, T. Francart, and A. Bertrand, “Auditory-inspired speech envelope extraction methods for improved EEG-based auditory attention detection in a cocktail party scenario,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 5, pp. 402–412, 2017.
  - [11] T. W. P. Janssen, M. Bink, K. Geladé, R. van Mourik, A. Maras, and J. Oosterlaan, “A randomized controlled trial into the effects of neurofeedback, methylphenidate, and physical activity on EEG power spectra in children with ADHD,” *Journal of Child Psychology and Psychiatry*, vol. 57, no. 5, pp. 633–644, 2016.
  - [12] V. Goverdovsky, D. Looney, P. Kidmose, and D. P. Mandic, “In-ear EEG from viscoelastic generic earpieces: robust and unobtrusive 24/7 monitoring,” *IEEE Sensors Journal*, vol. 16, no. 1, pp. 271–277, 2016.
  - [13] S. V. Eyndhoven, T. Francart, and A. Bertrand, “EEG-informed attended speaker extraction from recorded speech mixtures with application in neuro-steered hearing prostheses,” *IEEE Transactions on Bio-Medical Engineering*, vol. 64, no. 5, pp. 1045–1056, 2016.
  - [14] L. He, D. Hu, and M. Wan, “Common bayesian network for classification of EEG-based multiclass motor imagery BCI,” *IEEE Transactions on Systems Man & Cybernetics Systems*, vol. 46, no. 6, pp. 843–854, 2017.
  - [15] M. M. Stecker, D. Sabau, L. R. Sullivan et al., “American clinical neurophysiology society guideline 6: minimum technical standards for EEG recording in suspected cerebral death,” *The Neurodiagnostic Journal*, vol. 56, no. 4, pp. 276–284, 2016.
  - [16] J. Hofmeijer and M. J. A. M. Van Putten, “EEG in postanoxic coma: prognostic and diagnostic value,” *Clinical Neurophysiology*, vol. 127, no. 4, pp. 2047–2055, 2016.
  - [17] T. L. T. Da Silveira, A. J. Kozakevicius, and C. R. Rodrigues, “Single-channel EEG sleep stage classification based on a streamlined set of statistical features in wavelet domain,” *Medical & Biological Engineering & Computing*, vol. 55, no. 2, pp. 1–10, 2017.
  - [18] R. Gupta, K. ur Rehman Laghari, and T. H. Falk, “Relevance vector classifier decision fusion and EEG graph-theoretic features for automatic affective state characterization,” *Neurocomputing*, vol. 174, pp. 875–884, 2016.
  - [19] F. Qi, Y. Li, and W. Wu, “RSTFC: a novel algorithm for spatio-temporal filtering and classification of single-trial eeg,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 12, pp. 3070–3082, 2017.
  - [20] S. Finnigan, A. Wong, and S. Read, “Defining abnormal slow EEG activity in acute ischaemic stroke: delta/alpha ratio as an optimal QEEG index,” *Clinical Neurophysiology*, vol. 127, no. 2, pp. 1452–1459, 2016.
  - [21] A. Thul, J. Lechinger, J. Donis et al., “EEG entropy measures indicate decrease of cortical information processing in disorders of consciousness,” *Clinical Neurophysiology*, vol. 127, no. 2, pp. 1419–1427, 2016.
  - [22] J. GrNli, M. J. Rempe, and W. C. Clegern, “Beta EEG reflects sensory processing in active wakefulness and homeostatic sleep drive in quiet wakefulness,” *Journal of Sleep Research*, vol. 25, no. 3, pp. 257–268, 2016.
  - [23] M. Leitinger, E. Trinka, E. Gardella et al., “Diagnostic accuracy of the salzburg EEG criteria for non-convulsive status epilepticus: a retrospective study,” *The Lancet Neurology*, vol. 15, no. 10, pp. 1054–1062, 2016.
  - [24] G. Dippel, W. Chmielewski, and M. Muckschel, “Response mode-dependent differences in neurofunctional networks during response inhibition: an EEG-beamforming study,” *Brain Structure & Function*, vol. 221, no. 8, pp. 1–11, 2016.
  - [25] S. Charbonnier, R. N. Roy, S. Bonnet, and A. Campagne, “EEG index for control operators’ mental fatigue monitoring using interactions between brain regions,” *Expert Systems with Applications*, vol. 52, pp. 91–98, 2016.
  - [26] R. Zink, Hunyadi, Borbála, and S. V. Huffel, “Mobile EEG on the bike: disentangling attentional and physical contributions to auditory attention tasks,” *Journal of Neural Engineering*, vol. 13, no. 4, pp. 46017–46019, 2016.
  - [27] C. Cheng, X. Wei, and Z. Jian, “Emotion recognition algorithm based on convolution neural network,” in *Proceedings of 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, IEEE, Kuala Lumpur, Malaysia, November 2017.
  - [28] D. G. Dessavre, J. E. Ramirez-Marquez, and K. Barker, “Multidimensional approach to complex system resilience analysis,” *Reliability Engineering and System Safety*, vol. 149, no. 5, pp. 34–43, 2016.

## Research Article

# Collaborative Intelligent Environment Perception and Mission Control of Scientific Researchers in Semantic Knowledge Framework Based on Complex Theory

Jingfeng Zhao and Yan Li 

*College of Management and Economic, North China University of Water Resources and Electric Power, Zhengzhou 450000, Henan, China*

Correspondence should be addressed to Yan Li; [b2018120118@stu.ncwu.edu.cn](mailto:b2018120118@stu.ncwu.edu.cn)

Received 23 October 2020; Revised 20 November 2020; Accepted 11 December 2020; Published 30 December 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Jingfeng Zhao and Yan Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the traditional scientific research and production activities, due to the lack of sufficient communication and communication between researchers, the phenomenon of waste of scientific research resources occurs from time to time, which hinders the efficiency of scientific research output. Based on the design principle of the semantic knowledge framework, this paper puts forward the definition of ontology and semantic relationship of the collaborative system of scientific researchers. In this paper, a framework of collaborative semantic knowledge among researchers is established through decentralized semantic information exchange architecture. In this article, the simulation is verified by experiments and compared with other exchange architectures. The results of the experiment confirmed the semantic information exchange architecture based on semantic knowledge proposed in this paper is 10.39% faster than the traditional centralized method in terms of data volume; the construction speed under the data node perspective is 12.84% higher than that of the traditional centralized construction method; the subject query speed is 36.84% higher than that of the traditional centralization method; the predicate query speed is 31.58% higher than that of the traditional centralization method. The experimental results confirm that the semantic information exchange architecture based on the semantic knowledge framework is feasible, and it has excellent performance in terms of construction speed and query speed. Under the background that researchers rely more and more on collaborative technology to interact with other members, this paper has a certain reference value and exploration value and proposes a new idea of group collaboration system under the framework of semantic knowledge.

## 1. Introduction

With the advent of the information age, especially the development of web technology and computer-supported collaborative technology, the academic environment of modern science and technology has changed. The interdisciplinary and interpenetration between disciplines is a remarkable feature of modern science and also the main direction of scientific research. Due to the complexity and specialization of scientific research tasks, it is more difficult for researchers to complete scientific research tasks relying on personal knowledge reserves [1]. The tasks in the field of scientific research should be based on team cooperation. As a

high-efficiency organization form of scientific research, collaborative research will become an effective way to improve scientific research efficiency, reduce research costs, strengthen academic exchanges, and stimulate academic innovation. At the same time, it hinders the rapid development of information for users. As a special kind of users, researchers' information behavior is different from other information users. The essence of scientific research personnel is not to obtain a large amount of information but to obtain information that can solve problems. In traditional scientific research and production activities, due to the lack of sufficient communication and communication between researchers, the phenomenon of waste of scientific research

resources occurs from time to time, which hinders the efficiency of scientific research output. In terms of an isolated scientific research environment, collaborative scientific research cooperation has become necessary [2].

Allen et al.'s research shows that collaborative research has become the pillar of knowledge production in many scientific fields and has been promoted as a method to improve the quality of scientific research, resource utilization, and influence. But he did not do more in-depth research, leading to the conclusion is not comprehensive [3]. In terms of the motivation of collaborative research, Botsford et al. analyzed the difference between the experimental research and the theoretical research and concluded that the difference of professional division was the main factor that promoted the collaborative research. In the process of research projects, it is easier to obtain knowledge and experience of new partners than to master new knowledge by themselves. Knowledge resources obtained from collaborative partners can improve scientific research efficiency and output. Due to the lack of comparison objects selected in the research project, the conclusion cannot be used as a reference basis [4]. At the same time, scholars have also noticed the impact of scenario factors on collaborative research and development. Berbegal's impact on collaborative research and cooperation shows that regional, cultural, economic, and political factors are the main factors affecting collaborative research, while the degree of collaboration decreases with the increase of space and geography among research partners. However, he did not list the detailed data, resulting in his research conclusion is not rigorous [5]. In addition, transportation and communication technology is also one of the factors affecting collaborative research. The convenient traffic conditions and convenient communication technology eliminate the collaborative obstacles brought by the region and also greatly reduce the cost of scientific research, which provides favorable conditions for the collaborative cooperation and academic exchange between scientific research studies.

This research breaks through the pure theory behavior discussion, based on the theoretical analysis, supplemented by the investigation and analysis of practical application, combines the theory and practice, focuses on the collaborative information behavior of users in the collaborative research environment, and takes the collaborative research environment as the background, based on the semantic knowledge framework, the language of the environment perception system, the collaborative architecture, and the knowledge framework of the researchers. Finally, the feasibility of the semantic knowledge framework is verified by simulation experiments.

## 2. Theoretical Basis of Group Collaboration Architecture, Intelligent Environment Awareness, and Semantic Knowledge Framework

### 2.1. Intelligent Environment Awareness

**2.1.1. Convolution Neural Network.** CNN is an important model in-depth algorithm. It is a nonlinear and adaptive mathematical model for processing information, and it

consists of a large number of independent computing nodes. It is suitable for processing data with grid structure (such as image composed of two-dimensional pixel grid). CNN structure is similar to ANN, which is mainly divided into three parts: the input layer receives the signals and data outside the model and receives the original image information as input in CNN; the hidden layer performs nonlinear mapping on the input data to form the characteristics of different levels of data; and the output layer outputs discrete or continuous data processing results [6, 7]. Based on ANN, CNN introduces sparse connection, weight sharing, and downsampling technology to realize hierarchical processing of visual information.

As shown in Figure 1, CNN model has a variety of structures, including convolution layer, downsampling layer, and activation layer. Convolution layer, pooling layer, and activation layer are three basic structures to form function modules to realize feature coding and nonlinear mapping. Multiple functional modules form the deep model to realize the abstract expression of features. Finally, class labels and probability are output through full connection layer and loss layer [8, 9].

(1) *Convolution and Activation.* Convolution operation can be regarded as a linear operation process of weighted sum of two-dimensional images. At this time, the weight matrix used is called convolution kernel. Unlike the dense connection of neurons in ANN, the convolution kernel of CNN is associated with a certain region in the two-dimensional image, and the value after weighted sum is activated as the corresponding pixel value of the new feature map [10]. The convolution operation in CNN is as follows:

$$x_j^l = f \left( \sum_{i \in M_i} x_i^{l-1} * k_{ij}^l + b_j^l \right), \quad (1)$$

where  $f$  is the layer index;  $x$  is the feature graph, which is a two-dimensional matrix;  $i$  and  $j$  represent the input and output feature map indexes, respectively. Specifically,  $x_j^l$  is the  $j$ -th output characteristic graph of this layer;  $x_i^{l-1} * k_{ij}^l$  is the convolution operation of the output characteristic graph of the upper layer through kernel  $k_{ij}^l$ ;  $i \in M_i$  is the traversal operation of all input characteristic graph;  $*$  is the convolution operation;  $b_j^l$  is the offset direction; and  $f$  is the activation function.

The main function of the activation layer is to realize the nonlinear transformation of CNN. Activation functions have many variations, such as sigmoid function, tanh function, and relu function. Currently, relu activation function is commonly used, which has the advantages of stable error propagation for different input sizes and avoiding gradient explosion or dispersion. Moreover, the activation function has zero response to negative input and can realize the sparse connection of the network [11].

(2) *Pooling.* Pooling layer is also known as the lower sampling layer, and the common ones are Max pooling, average pooling, and min pooling. The pooling layer can reduce the spatial resolution of the feature map, so as to reduce the

network scale, accelerate the network training, reduce the overfitting, and make the features have strong translation and scaling invariance for different inputs [12]. The general expression of pooling operation is as follows:

$$x_j^l = f(\beta_j^l \text{down}(x_j^{l-1}) + b_j^l). \quad (2)$$

Down( $\cdot$ ) is the partition calculation operation of the feature graph, such as dividing the feature graph into  $n \times n$  grids and calculating the sum, maximum value, and minimum value of each part.  $\beta_j^l$  is the weight parameter of grid elements; 1 is taken for maximum pooling;  $1/S$  is taken for average pooling ( $s = w \times h$ , where  $w$  and  $h$  are pool kernel sizes, respectively); and  $f$  is the activation function.

**2.1.2. MobileNet.** MobileNet is a lightweight CNN model designed for embedded hardware platform. By introducing a depth separable convolution layer, the standard convolution is decomposed into a combination of depth convolution which only extracts features from a single channel and point convolution that fuses all channel information, thus greatly reducing the amount of parameters and realizing model acceleration [13]. The number of input channels is  $n_i$ , and the size of the input characteristic graph is  $w_i$  and  $h_i$ . The corresponding output channel number and characteristic layer sizes are  $n_{i+1}$ , and  $w_{i+1}$  and  $h_{i+1}$ , respectively. If the size of the convolution kernel is  $k_i \times k_i \times n_i$ , the number of standard convolution multiplication operations is given by

$$k_i \times k_i \times n_i \times n_{i+1} \times h_i \times w_i. \quad (3)$$

Deep convolution is a special case in which the number of blocks is equal to the number of channels in block convolution:

$$k_i \times k_i \times n_i \times h_i \times w_i. \quad (4)$$

Deep convolution is independent in different channels, so there is a problem that information does not flow between channels. In MobileNet, different channel information is combined in the form of point convolution between channels:

$$1 \times 1 \times n_i \times n_{i+1} \times h_i \times w_i. \quad (5)$$

When the standard convolution is replaced by separable convolution, the compression ratio of the model is given by

$$\frac{k_i \times k_i \times n_i \times h_i \times w_i + n_i \times n_{i+1} \times h_i \times w_i}{k_i \times k_i \times n_i \times n_{i+1} \times h_i \times w_i} = \frac{1}{n_{i+1}} + \frac{1}{k_i^2}. \quad (6)$$

Generally, the convolution kernel size of CNN structure is  $3 \times 3$ , and the number of output characteristic layers  $n_{i+1}$  is often large. According to formula (6), when the standard convolution operation is replaced by separable convolution, the compression ratio of model parameters and multiplication operation is  $1/8 \sim 1/9$ .

The design of the MobileNet grid structure is very similar to VGGNet. The spatial resolution of the feature map decreases monotonously, and the number of channels increases monotonously. When the resolution is reduced by

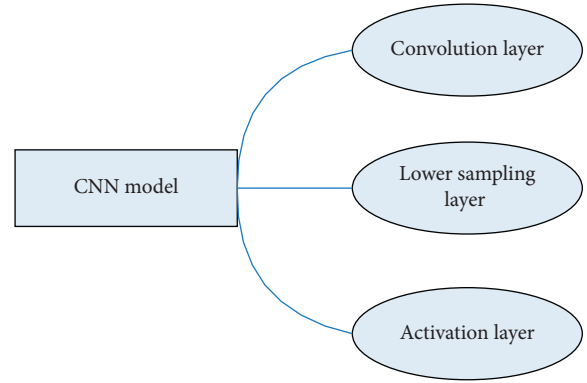


FIGURE 1: CNN mode structure.

half, the number of feature layers is doubled. The differences are as follows:

- (1) MobileNet replaces the standard convolution with deep separable convolution in VGGNet
- (2) MobileNet does not set the pooling layer but reduces the spatial resolution of the feature map by the convolution step size of 2
- (3) MobileNet replaces the two full link layers of VGGNet  $4096 \times 4096$  and  $4096 \times 1000$  with  $7 \times 7$  average pooling layer, realizing further compression of the model [14]

## 2.2. Principle Design of Semantic Knowledge Framework

**2.2.1. Semantic Knowledge Framework.** Semantic knowledge framework can be considered as an analysis method in philosophy. The analysis method has one basic element and two cores. The basic element is concept. The first core is the relationship between concepts. The second core is to realize reasoning function through the relationship between concepts. Second, the semantic knowledge framework can be a form of computer knowledge storage and representation and concept. Finally, the semantic knowledge framework can also be used for knowledge processing, such as identification, reasoning, query, knowledge consistency maintenance, scenario calculus, and planning so as to achieve knowledge fusion, knowledge extraction, knowledge discovery, natural language generation, and other functions [15, 16].

Semantic knowledge framework can be used to describe any complex relationship between anything, but this description is based on a series of basic semantic relations. Basic semantic relations are the basic elements of complex semantic knowledge framework. The basic semantic relations are various and flexible. As shown in Figure 2, here are some commonly used relations.

(1) *Semantic Relation.* Semantic relation generally describes the generic relationship between things, including IS-A, A-Kind-Of, and Instance-Of.

IS-A means that one thing is an instance of another. It can be expressed as “. . . It’s an example of.” For example, if a geological researcher is regarded as a class and a geological

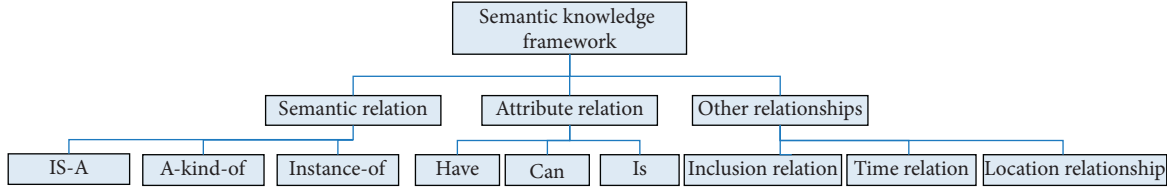


FIGURE 2: Framework diagram of the semantic knowledge framework.

exploration technician is a member of the class, then the geological exploration technician is an example of a geological researcher.

A-Kind-Of means that one thing is a type of another “It’s a kind of . . .” AKO represents a larger range than IS-A. It usually does not represent the relationship between specific individuals but represents the relationship between classes. AKO is generally used to establish the relationship between subclasses and superclasses. For example, geological researchers are a subclass, scientific researchers are a parent class, and the semantic relationship between geological researchers and researchers can be represented by AKO.

The Instance-Of relation is the opposite of IS-A relation, which means that one instance of something is another.

(2) *Attribute Relation.* Attribute relationship usually refers to the relationship between things and their attributes. Any object in any class has one or more properties, and each property corresponds to a value. Therefore, there will be a corresponding combination of properties and values. The commonly used attribute relation is the predicate or verb of a sentence. For example, Have, Can, and Is.

Have means that things have a certain attribute relationship, which is expressed as “have.”

Can means the relationship between certain attributes of things, which can be expressed as “can” or “will.”

Is has no specific representation and can be understood as a variety of relationships. If something has multiple relationships with other things or attributes, they can be connected through is.

(3) *Other Relationships.* The relationship in the real world is very complex. In addition to the above semantic relations and attribute relations, there are many kinds of relationships between things and between things and attributes. The other main relationships are inclusion relationship, time relationship, location relationship, and similarity relationship [17].

The inclusion relation represents the relationship between the whole and the part. The difference between inclusion relation and attribute relation is that inclusion relation can be inherited and part of it belongs to whole, but it has all attributes of whole. Inclusion relationships can be described as Part-of or Composed-of.

The time relation represents the sequence of events in time. For example, Before means that an event must occur before a specific event occurs; At means that an event occurs at the same time as another event; and After means that an event can only occur after a specific event occurs.

Positional relations represent the relationship between things in space. If the position of one thing is in front of another, it can be represented by Location-front; if something is behind another, it can be expressed by Location-behind.

2.2.2. *Knowledge Graph.* Knowledge map is actually a semantic network, which can form reasoning semantic knowledge network from the connection of different semantic entities according to the change (relationship). The representation form is a graphical structure. In fact, the construction process of knowledge map includes the following: integrating the form of semantic knowledge and data cleaning form into heterogeneous data sources, establishing the relationship model through relation extraction, and finally establishing a directed graph structure database, which can reflect the semantic relationship between entities [18]. The query of the knowledge map is based on visual query. The knowledge obtained through input information is not a large number of web pages obtained by string matching, but the structured knowledge that users really need.

### 2.3. Swarm Intelligence Algorithm

#### 2.3.1. Cuckoo Search Algorithm

(1) *Lévy Flight.* Lévy distribution is a kind of continuous probability distribution. The letters  $\delta$ ,  $\alpha$ ,  $\mu$ , and  $\beta$  represent the scale, characteristic index, displacement, and skewness parameters, respectively:

$$\begin{aligned}
 P_{\alpha,\beta}(\kappa; \mu, \sigma) &= F\{P_{\alpha,\beta}(x; \mu, \sigma)\} = \int_{-\infty}^{\infty} dx e^{ikx} P_{\alpha,\beta}(x; \mu, \sigma) \\
 &= \exp\left[iuk - \sigma^\alpha |k|^\alpha \left(1 - i\beta \frac{k}{|k|} \bar{\omega}(k, \alpha)\right)\right], \\
 \bar{\omega}(k, \alpha) &= \begin{cases} \tan \frac{\pi\alpha}{2}, & \text{if } \alpha \neq 1, 0 < \alpha < 2, \\ -\frac{2}{\pi} \ln|k|, & \text{if } \alpha = 1. \end{cases}
 \end{aligned} \tag{7}$$

The probability density function of the Lévy distribution is related to  $\alpha$  characteristic index and  $\beta$  skewness parameter. When  $\alpha$  and  $\beta$  take different values, they can be expressed by different distribution functions (such as Gaussian distribution, Cauchy distribution, and Lévy distribution).



When  $\alpha = 2$ , the following Gaussian distribution function is used:

$$p_2(x) = \frac{1}{\sqrt{4\pi}} \exp\left(-\frac{x^2}{4}\right). \quad (8)$$

When  $\alpha = 1$  and  $\beta = 0$ , it is expressed by the Cauchy distribution function:

$$p_{1,0}(x) = \frac{1}{\pi(1+x^2)}. \quad (9)$$

When  $\alpha = 1/2$  and  $\beta = 1$ , it is expressed by the Lévy distribution function:

$$p_{1/2,1}(x) = \begin{cases} \frac{1}{\sqrt{2\pi}} x^{-(3/2)} \exp\left(-\frac{1}{2x}\right), & x \geq 0, \\ 0, & x < 0. \end{cases} \quad (10)$$

The waiting time of Lévy flight jump length obeys the power law distribution function:

$$\text{Lévy}(\beta) \sim \mu = t^{-1-\beta}, \quad 0 < \beta \leq 2. \quad (11)$$

There are two main elements in Lévy flight, which are moving direction and jumping step size  $s$ . In Mantegana's law, the definition of step length  $s$  is as follows:

$$s = \frac{\mu \times \varepsilon}{|\nu|^{1/\beta}}, \quad (12)$$

where  $\mu$  and  $\nu$  obey the standard normal distribution,  $\beta = 1.5$ .

$$\varepsilon = \left( \frac{\Gamma(1+\beta) \times \sin(\pi \times \beta/2)}{\Gamma((1+\beta)/2) \times \beta \times 2^{(\beta-1)/2}} \right)^{1/\beta}. \quad (13)$$

**2.3.2. Clustering Algorithm.** In K-means algorithm, the sum of squares of errors is used as the partition criterion:

$$E(C) = \sum_{i=1}^K \sum_{x_j \in C_i} \|x_j - \mu_i\|^2. \quad (14)$$

K-medoids clustering algorithm uses the actual data points as the clustering center and takes the absolute error as the division:

$$E = \sum_{j=1}^k \sum_{x \in N_j} |x - C_j|. \quad (15)$$

**2.3.3. Fast Search Algorithm of Density Peak.** For each data point  $I$ , the DPC algorithm needs to calculate its local density  $\rho_i$  and its distance  $\delta_i$ . When the set of data points is large, the local density  $\rho_i$  of data point  $i$  is calculated by the following formula:

$$\rho_i = \sum \chi(d_{ij} - d_c), \quad (16)$$

where  $X(x)$  is given by

$$\chi(x) = \begin{cases} 1, & x < 0, \\ 0, & x > 0. \end{cases} \quad (17)$$

When the set of data points is small, the local density of data points is calculated by the exponential kernel:

$$\rho_i = \sum_j e^{-(d_{ij}/d_c)^2}. \quad (18)$$

The formula for calculating the distance of data point  $I$  is as follows:

$$\delta_i = \min_{j: \rho_j > \rho_i} (d_{ij}). \quad (19)$$

For local density maximum data point  $i$ , the distance formula is as follows:

$$\delta_i = \max_j (d_{ij}). \quad (20)$$

### 3. Design of Simulation Experiment

#### 3.1. Group Collaboration Architecture

(1) Introduction: the design of group architecture is related to the level and function division of each unit in the system and is the basis of each unit structure design [19]. The structure of the unit determines the assignment of tasks, the content of information flow, and the specific stage of task execution. The design of group architecture should follow the following principles:

- (a) Clear hierarchy: hierarchical relationship includes two parts: group level and internal operation level. A clear hierarchical relationship helps to plan and integrate each unit into the system independently; the hierarchical structure of internal functional units is conducive to the standardization and design of the system and is convenient to adjust and expand the system [20, 21].
- (b) Reasonable function distribution: it is necessary to make logical planning for the functions of different levels in the system, so as to avoid some functions being too complex and others relatively single. Otherwise, it will not only affect the overall performance of the system but also reduce the execution efficiency due to the large amount of data processing, leading to the paralysis of some system units.
- (c) Efficient information transmission: the content and form of information transmission between units and operation modules should be fully considered in the design of system structure. In the unit, the standardization of information and the design of the summary process are of great significance to improve the execution efficiency and reduce the cost of the system.

- (2) As shown in Figure 3, the organizational forms of group collaboration can be divided into centralized, decentralized, and distributed.

**3.1.1. Centralized Control Structure System.** The system is controlled by a main control unit, which is a top-down hierarchical control structure for planning and decision-making. The number and complexity of the main control units determine the time required to respond to the system and the quality of decision-making behavior [22]. The main control unit is responsible for the dynamic allocation of tasks and the potential planning of resources and coordinates the competition and cooperation among various posts. The system is easy to manage, control, and program.

**3.1.2. Decentralized Control Structure System.** Each individual in the system has an equal relationship, has a high degree of intelligent autonomy, independently processes information, design, and decision-making, performs its own tasks, and communicates with other units to coordinate their behavior without a central control unit [23]. The structure has good flexibility, scalability, and reliability, but the communication requirements are high and the efficiency of multilateral negotiation is low. Therefore, it is difficult or unable to guarantee the realization of the global goal.

**3.1.3. Distributed Control Structure System.** This structure is the product of the combination of decentralized horizontal interaction and centralized vertical control. It is composed of a group of independent, completely equal, no logical master-slave relationship and self-discipline. According to the predefined protocol, according to the system goal, state, and its own state, ability, resources, and knowledge, each unit uses the communication network to consult and negotiate with each other to determine their respective tasks, coordinate their own activities, realize the sharing of resources, knowledge, information, and functions, and cooperate to complete common tasks to achieve the overall goal. In this kind of system, each unit is independent of each other in structure and function, and they all communicate with each other through the network in the same way, with good encapsulation, so the system has good fault tolerance, openness, and scalability [24, 25].

**3.2. Experimental Dataset.** In this paper, we build a decentralized system of 6 nodes in win10 environment. The protocol includes Hypertext Transfer Protocol and Peer-to-Peer. The ArchiveHub dataset is used in the experiment. The size of the dataset is 72 m, the quantity of the substance is 107219, the number of subjects is 51385, the number of unique predicates is 143, the quantity of unique objects is 104389, and the number of triples is 432142. In this article, the data collection is carved up into six parts, named X1~X6, construct a respective knowledge map for each data collection and save it in six nodes.

Table 1 shows the essential information of the six datasets. In the subsequent comparative experiments, the trend of construction time and query time is related to the trend of the number of unique entities. It can more directly reflect the series ring in the process of physical connection building the global knowledge map, and a copy of the selected node information is copied in XL-X6.

**3.3. Verification Test.** Building module verification experiment: Several entity modules are randomly selected for construction. In this process, we can explore and verify the construction effect of the semantic information exchange system by viewing the connection period between the entity modules. With physical resources self3810 as an example, self3810 points to dataset X2 (port number 8002) in the connection information table of the node where dataset X1 (port number is 8001) is located and in the connection information table of the node where dataset X2 (port number 8002) points to dataset X3 (port number 8003). Dataset X3 (connection information table with port number 80031) points to dataset X1 (port number 8001). From the above results, the entity resource self3810 forms a link cycle in the three nodes of the system, and the link cycle trend is as follows: the next largest node is the next node pointed to by the current node.

## 4. Comparison of Test Results of Different Methods

This article mainly conducts experiments from two aspects: the construction speed of the semantic knowledge framework and the query speed.

**4.1. Construction Rate Comparison (Data Volume Perspective).** The contrast experiment design of construction rate is compared with the centralized construction system from two dimensions of data size and node number size.

As shown in Figure 4 and Table 2, from the perspective of data volume, the construction rate of other methods and the semantic knowledge framework system proposed in this paper is compared. Taking three nodes as a group, the amount of data gradually increases to observe the respective performance of the centralized system and decentralized system. It can be concluded from the figure that

- (1) With the increase of data volume, the construction time of both centralized and decentralized knowledge maps will show a slow upward trend. A very important part of the construction process of the knowledge map is the data transmission and connection, and the construction time is mainly increased in the data transmission connection part. Therefore, because of the growth of data volume, the construction time of both the centralized system and the decentralized system will increase correspondingly.

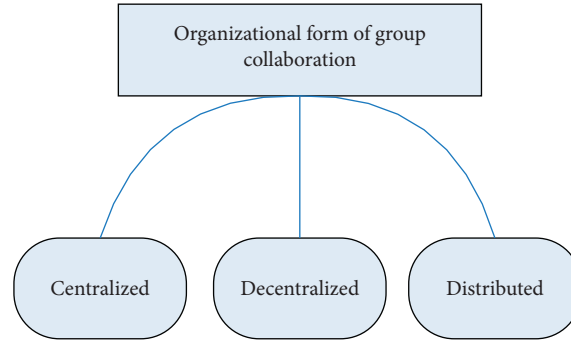


FIGURE 3: Organization form chart of group collaboration.

TABLE 1: Basic information of experimental data collection.

| Data collection | File size | Quantity of unique substance | Quantity of triplet | Quantity of unique subjects | Unique predicate quantity |
|-----------------|-----------|------------------------------|---------------------|-----------------------------|---------------------------|
| X1              | 5.8       | 16875                        | 29746               | 6971                        | 14082                     |
| X2              | 6.9       | 21538                        | 39910               | 9285                        | 13369                     |
| X3              | 9.4       | 20364                        | 54189               | 2614                        | 20871                     |
| X4              | 15.7      | 29483                        | 79852               | 4536                        | 27259                     |
| X5              | 16.1      | 40576                        | 101987              | 12857                       | 30364                     |
| X6              | 18.9      | 43295                        | 123948              | 14921                       | 30428                     |

(2) According to Figure 1, when the size of the dataset is small, the construction speed of the proposed semantic knowledge framework system is lower than that of the traditional construction method. But, with the increase of datasets, the semantic knowledge framework system gradually shows its advantages. The reason is that the system constructed by the traditional method needs to transmit all datasets, but in order to protect the information of each node, the system only transmits the unique entity set. The reason why the rate of early construction of the system proposed in this paper is slower than that of the traditional method is that the empty node needs to send the generated information back to each node, which takes a lot of time.

#### 4.2. Construction Rate Comparison (Data Node Perspective).

As shown in Table 3 and Figure 5, this paper compares the construction rate of traditional methods with that of the semantic knowledge framework system proposed in this paper. With the increasing number of nodes, the performance of the centralized system and decentralized system is observed. It can be concluded from the figure that

- (1) From the overall trend, with the increase in the number of nodes, the construction time of both centralized and decentralized semantic knowledge framework will show an upward trend. Since the construction of the semantic knowledge framework requires the connection between nodes and data transmission, with the increase in the number of nodes, the construction time of both centralized and decentralized systems will increase correspondingly.
- (2) As shown in Figure 5, we can see that the construction rate of the semantic information exchange

architecture in this paper is lower than that of the traditional semantic information exchange architecture in the case of a small number of nodes. However, with the increasing number of nodes, the construction speed of the decentralized semantic information exchange architecture proposed in this paper is gradually higher than that of the traditional semantic information exchange architecture construction. The reason is that the traditional method of transmitting information only needs to protect part of the dataset. The reason for the slow construction speed of the system proposed in this paper is that the empty node needs to return the generated connection information to each node, which takes part of the time.

#### 4.3. Comparison of the Number of Subject-Predicate Word Number Queries.

As shown in Table 4 and Figure 6, the query time of the two types of systems under different subjects is shown.

As shown in Table 5 and Figure 7, the query time of the two systems under different predicates is shown. It can be concluded from Figure 4 that

- (1) With the increasing number of subject-predicate words, the query time will also increase. This is because the number of query levels in the query mode is determined by the number of subject-predicate words. If the number of subject-predicate words increases, the number of query connection levels will increase.
- (2) In terms of the same subject or predicate word number, the query speed of the decentralized query mode is faster than that of the centralized query mode. This is because in the decentralized query

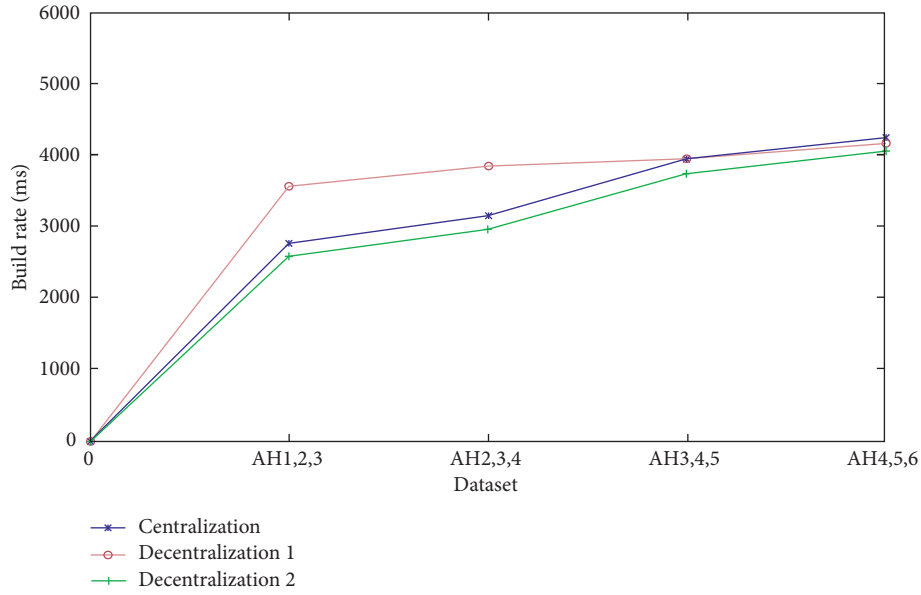


FIGURE 4: Rate comparison (data volume).

TABLE 2: Rate comparison (data volume).

|               | AH1,2,3 | AH2,3,4 | AH3,4,5 | AH4,5,6 |
|---------------|---------|---------|---------|---------|
| Decentralized | 2784    | 3129    | 3926    | 4089    |
| Centralized 1 | 3632    | 3775    | 3851    | 3967    |
| Centralized 2 | 2564    | 2941    | 3472    | 3845    |

TABLE 3: Rate comparison (data node).

| Build time    | 1    | 2    | 3    | 4    | 5    |
|---------------|------|------|------|------|------|
| Decentralized | 516  | 524  | 1418 | 1895 | 2654 |
| Centralized 1 | 1136 | 957  | 1594 | 2015 | 1720 |
| Centralized 2 | 1306 | 1274 | 1453 | 1801 | 1651 |

TABLE 4: Query rate comparison (subject).

| Query time    | 1   | 2   | 3   | 4   |
|---------------|-----|-----|-----|-----|
| Decentralized | 58  | 102 | 167 | 194 |
| Centralized 1 | 103 | 181 | 253 | 328 |
| Centralized 2 | 109 | 196 | 261 | 341 |

mode, queries are parallel, and the connection between tables is based on the connection information part of each node dataset; in the centralized system, the table join during query is the self-join of the whole large dataset.

- (3) In terms of the overall trend, the increase of query time in the decentralized query mode is gradually decreasing, while that in the centralized query mode is unchanged. This shows that with the increase of subject-predicate words, the advantage of the decentralized query mode will become more and

more obvious. This is because each table join of a centralized query queries the semantic knowledge framework, so the increase is not changed, and each connection of the distributed query is based on the connection information part of each node dataset. In summary, the performance of decentralized semantic information exchange query mode based on the semantic knowledge framework of group collaborative intelligent environment and mission control is better than that of the traditional centralized query mode.

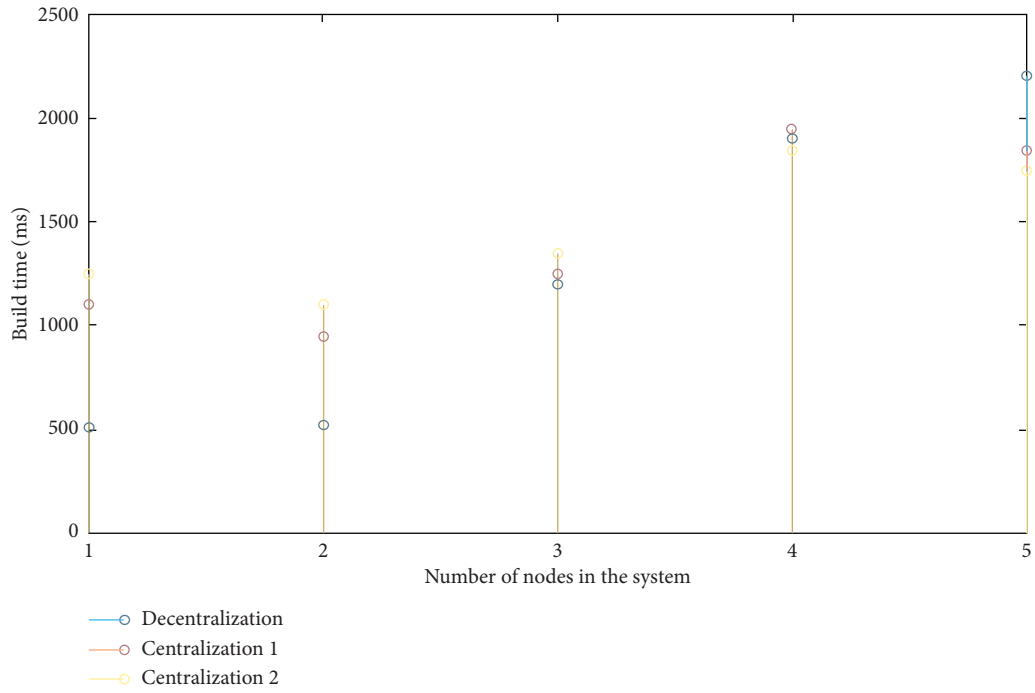


FIGURE 5: Rate comparison (data node).

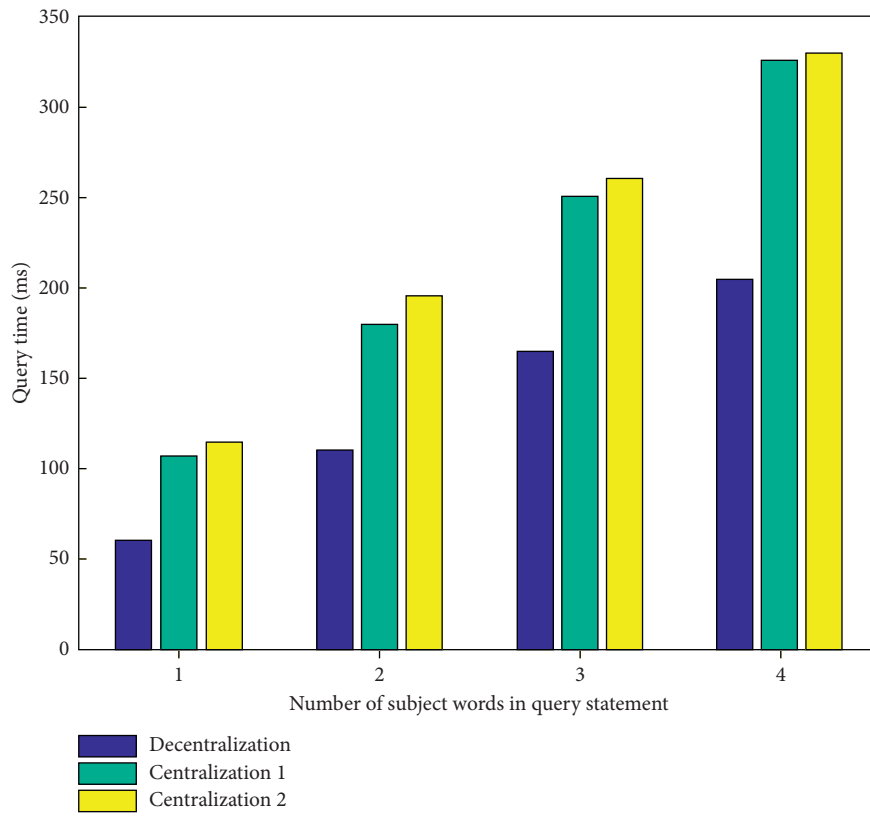


FIGURE 6: Query rate comparison (subject).

TABLE 5: Query rate comparison (predicate).

| Query time    | 1   | 2   | 3   | 4   |
|---------------|-----|-----|-----|-----|
| Decentralized | 51  | 105 | 154 | 187 |
| Centralized 1 | 101 | 194 | 246 | 337 |
| Centralized 2 | 107 | 203 | 259 | 352 |

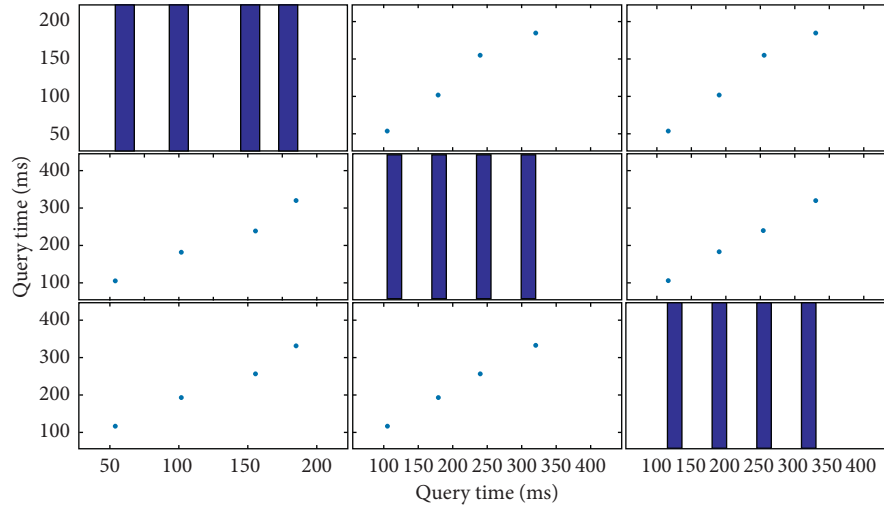


FIGURE 7: Query rate comparison (predicate).

## 5. Conclusions

The main research content of this paper is based on the semantic knowledge framework of collaborative intelligent environment perception and mission control. In this paper, the concept of environmental awareness system, group collaboration, and semantic knowledge framework is introduced and analyzed in detail. In this paper, through the establishment of temporary space-time nodes with the same level as other nodes in the network, the interconnection and interaction between nodes can be realized, and the self-determination mechanism can realize the knowledge connection between nodes on the basis of maintaining their own knowledge. On the basis of the interaction mechanism of Unicom, we design and implement the decentralized Iterative Incremental Construction Scheme of the semantic knowledge framework and the corresponding query mode. On the premise that knowledge is not acquired, the connection construction and query between nodes are realized.

Experiments show that the semantic information exchange system structure constructed in this paper is feasible and effective. The integrity of the global semantic knowledge framework and the centralized semantic knowledge framework is the same, and the global semantic knowledge framework constructed in this paper has better performance than the middle school system in the construction speed and query rate.

At present, there are few research studies on the semantic knowledge framework, and there are still many deficiencies in this paper due to the limitation of time, specialty, and technical level, for example, lack of rule layer, only realize the sharing of data rather than rules; need to further improve the

algorithm to improve the query efficiency; and do not discuss the specific factors and influence coefficient of collaborative information in the collaborative research environment. The above shortcomings will be the next step of this project.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] M. Petera, G. Le Corquille, and M. Landi, "Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics," *Bioinformatics*, vol. 31, no. 9, pp. 1493–1495, 2015.
- [2] J. Vom Brooke and S. Lippe, "Managing collaborative research projects: a synthesis of project management literature and directives for future research," *International Journal of Project Management*, vol. 33, no. 5, pp. 1022–1039, 2015.
- [3] J. Allen, G. V. Mohatt, S. Beehler et al., "People awakening: collaborative research to develop cultural strategies for prevention in community intervention," *American Journal of Community Psychology*, vol. 54, no. 1-2, pp. 100–111, 2016.
- [4] J. Botsford, A. Ryman, J. Slonim et al., "Proceedings of the 1992 conference of the centre for advanced studies on collaborative research—volume 1," *Iranian Journal of Medical Education*, vol. 1, no. 10, pp. 851–859, 2015.
- [5] J. Berbegal-Mirabent and C. Llopis-Albert, "Applications of fuzzy logic for determining the driving forces in collaborative

- research contracts,” *Journal of Business Research*, vol. 69, no. 4, pp. 1446–1451, 2016.
- [6] D. Mishkin, N. Sergievskiy, and J. Matas, “Systematic evaluation of convolution neural network advances on the imagenet,” *Computer Vision and Image Understanding*, vol. 161, pp. 11–19, 2017.
- [7] A. Alotaibi and A. Mahmood, “Deep face liveness detection based on nonlinear diffusion using convolution neural network,” *Signal, Image and Video Processing*, vol. 11, no. 4, pp. 713–720, 2017.
- [8] P. Ghamisi, Y. Chen, and X. X. Zhu, “A self-improving convolution neural network for the classification of hyperspectral data,” *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 10, pp. 1537–1541, 2017.
- [9] Z. Liu, X. Yu, Y. Gao, S. Chen, X. Ji, and D. Wang, “CU partition mode decision for HEVC hardwired intra encoder using convolution neural network,” *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5088–5103, 2016.
- [10] L. Gao, P.-Y. Chen, and S. Yu, “Demonstration of convolution kernel operation on resistive cross-point array,” *IEEE Electron Device Letters*, vol. 37, no. 7, pp. 870–873, 2016.
- [11] S. Ur Rehman, S. Tu, Y. Huang et al., “CSFL: a novel unsupervised convolution neural network approach for visual pattern classification,” *AI Communications*, vol. 30, no. 5, pp. 1–14, 2017.
- [12] J. Gan, K. Jiang, H. Tan, and G. He, “Facial beauty prediction based on lighted deep convolution neural network with feature extraction strengthened,” *Chinese Journal of Electronics*, vol. 29, no. 2, pp. 312–321, 2020.
- [13] A. S. Alon, J. L. Dioses Jr., R. Diamante, R. M. Dellosa, and F. J. P. Montalbo, “MobileNet SSDv2 inference approach of smoke hazard detection and alert system: a smoke-induced simulated home-environment,” *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, pp. 197–202, 2020.
- [14] S. Purwar, R. K. Tripathi, R. Ranjan et al., “Detection of microcytic hypochromia using CBC and blood film features extracted from convolution neural network by different classifiers,” *Multimedia Tools and Applications*, vol. 79, no. 7–8, pp. 4573–4595, 2020.
- [15] N. Mishra, H. T. Chang, and C. C. Lin, “An IoT knowledge reengineering framework for semantic knowledge analytics for BI-services,” *Mathematical Problems in Engineering*, vol. 2015, Article ID 759428, 12 pages, 2015.
- [16] J. A. Nasir, I. Varlamis, and S. Ishfaq, “A knowledge-based semantic framework for query expansion,” *Information Processing & Management*, vol. 56, no. 5, pp. 1605–1617, 2019.
- [17] J.-R. Ruiz-Sarmiento, C. Galindo, and J. Gonzalez-Jimenez, “Exploiting semantic knowledge for robot object recognition,” *Knowledge-Based Systems*, vol. 86, pp. 131–142, 2015.
- [18] H. Paulheim and P. Cimiano, “Knowledge graph refinement: a survey of approaches and evaluation methods,” *Semantic Web*, vol. 8, no. 3, pp. 489–508, 2017.
- [19] A. Nitin, H. Ehtesham, L. Huan et al., “A subspace clustering framework for research group collaboration,” *International Journal of Information Technology & Web Engineering*, vol. 1, no. 1, pp. 35–58, 2017.
- [20] A. S. Messersmith, “Preparing students for 21st century teamwork: effective collaboration in the online group communication course,” *Communication Teacher*, vol. 29, no. 4, pp. 219–226, 2015.
- [21] X. Wang, C. Schneider, and J. S. Valacich, “Enhancing creativity in group collaboration: how performance targets and feedback shape perceptions and idea generation performance,” *Computers in Human Behavior*, vol. 42, pp. 187–195, 2015.
- [22] G. A. Jacquet, J. I. Schneider, J. Hudspeth et al., “A centralized structure and process to approve and monitor GME global health electives at a large academic institution,” *Journal of Graduate Medical Education*, vol. 9, no. 4, pp. 547–548, 2017.
- [23] H. Xin, R. Zhao, L. Zhang, Z. Wang, K. P. Wong, and W. Wei, “A decentralized hierarchical control structure and self-optimizing control strategy for F-P type DGs in islanded microgrids,” *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 3–5, 2016.
- [24] M. R. B. Khan, R. Jidin, and J. Pasupuleti, “Multi-agent based distributed control architecture for microgrid energy management and optimization,” *Energy Conversion & Management*, vol. 112, pp. 288–307, 2016.
- [25] G. Łukawski and K. Sapiecha, “Fault tolerant control for RP\* architecture of scalable distributed data structures,” *IEEE Transactions on Medical Imaging*, vol. 31, no. 4, pp. 870–881, 2015.

## Research Article

# Impact of Financial R&D Resource Allocation Efficiency Based on VR Technology and Machine Learning in Complex Systems on Total Factor Productivity

Hui Sun <sup>1</sup> and Xiong Zhong <sup>2</sup>

<sup>1</sup>School of Economics and Statistics, Guangzhou University, Guangzhou 510006, Guangdong, China

<sup>2</sup>Institute of Finance, Guangzhou University (Guangzhou Institute of International Finance), Guangzhou 510006, Guangdong, China

Correspondence should be addressed to Xiong Zhong; [zhongxiong@gzhu.edu.cn](mailto:zhongxiong@gzhu.edu.cn)

Received 17 October 2020; Revised 2 December 2020; Accepted 15 December 2020; Published 28 December 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Hui Sun and Xiong Zhong. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of the globalization of science and technology, innovation has become an important driving force for regional economic development. As a core element of regional innovation, financial R&D resources have also become a key element to enhance national innovation capabilities and national economic competitiveness. National and regional innovation capabilities have a direct impact. There are also many deep-seated problems behind the world-renowned achievements, such as irrational industrial structure, insufficient independent innovation capabilities, low resource utilization efficiency, and the service quality and efficiency of financial institutions for the transformation of total factor productivity. These problems extremely restrict the efficiency upgrade and further development of our country's total factor productivity. This study uses the DEA-Malmquist index model to measure the efficiency of fiscal R&D resource allocation in 28 provinces and regions in China in the past 10 years and uses Mapinfo12.0 software to analyze regional differences in the efficiency of fiscal R&D resource allocation in China from a spatial perspective. During the year, the overall R&D resource allocation efficiency of 28 provinces and autonomous regions in China has shown an upward trend. The efficiency of fiscal R&D resource allocation and the concentration of financial factors have had a positive impact on total factor productivity, transform and upgrade factors, increase total factor productivity, and provide empirical evidence for building a strong country.

## 1. Introduction

**1.1. Background and Significance.** In recent years, China has been steadily increasing R&D investment to improve its own level of innovation, but R&D varies greatly between different regions. Therefore, studying the efficiency of our country's R&D resource allocation and its regional differences is of great significance for reducing the differences in the efficiency of R&D resource allocation among regions, improving total factor productivity, enhancing regional innovation capabilities, and promoting the coordinated development of regional economies. As the main force of innovation resources, the effective allocation of R&D resources is playing an increasingly important role in the

process of economic development [1]. Total productivity and economic growth are the two main driving forces of modern economic growth and are important indicators of the quality of economic growth in a country or region. In order to cope with the long-term imbalance of our country's economic growth, the overall productivity level of each region has become an important basis for measuring economic efficiency and economic growth [2].

**1.2. Related Work Research.** Domestic research on the efficiency of fiscal R&D resource allocation is mainly carried out by different regions, different industries, and different R&D entities. The stochastic boundary analysis method is



used to measure the effectiveness of R&D resources in 28 provinces and regions. The factors affecting the effectiveness of regional R&D in our country from 2014 to 2019 are analyzed. DEA model and SFA model are mainly used to study the efficiency of R&D resource allocation in various regions of our country. The conclusions are basically the same. It can be considered that in the eastern, central, and western regions of our country, the allocation efficiency of R&D resources is significantly different. In regions with better economic growth, the allocation efficiency of R&D resources may not be high. The efficiency of allocating R&D resources to economically underdeveloped regions is not necessarily low, and DEA is effective or ineffective [3].

As early as the 1960s, scholars began to pay attention to the efficiency of R&D resource allocation, using the Cobb-Douglas production function to measure the efficiency of R&D resource allocation, using only R&D stock as an input indicator, select data from 40 countries and regions from 1960 to 2000 to study the effectiveness of R&D resource allocation in each country, and analyze its influencing factors. The survey results show that per capita GDP has the greatest impact on the efficiency of R&D resource allocation, and they also have a positive impact on the efficiency of R&D allocation [4, 5].

Liu G used nonparametric production frontier methods to study the components of our country's total factor productivity growth from two different perspectives: technological contribution and factor contribution. By combining two-year environmental technology and nonradiative directional distance functions, he developed a biennial. The nonradiation direction distance function measurement can overcome the well-known infeasibility problem, incorporate the total slack of the variable, and explore the driving factors of our country's TFP. Although this growth rate has slowed down recently, our country's TFP is still growing at an annual rate of 2.02%. Secondly, from the perspective of technological contribution, technological progress is the main driving force of total factor productivity [6]. Tugcu CT studied the causal relationship between energy consumption and TFP growth in the BRIC countries. Using panel bootstrapping Granger causality test, there is no significant causal relationship between renewable energy consumption and TFP growth in BRIC countries. Concerning nonrenewable energy, there is a two-way causal relationship between Brazil and South Africa's total factor productivity growth [7].

**1.3. Innovation.** Based on the New Economic Development Theory, Resource Allocation Theory, New Economic Geography Theory, etc., this article examines the effectiveness of our country's R&D resource allocation and its spatial differences from the perspective of time and space and combines them organically. Economy and geography make it more intuitive and clear to reflect the geographical and spatial arrangement of our country's R&D activities and their temporal and spatial evolution.

In this study, the DEA-Malmquist index model and factor analysis method are used to conduct an in-depth study

of the regional differences in the effectiveness of R&D resource allocation and environmental impact factors in our country. The effectiveness of R&D allocation is broken down into technical performance and pace of technological progress, as well as regions and regions. We conduct a more thorough and specific analysis of the differences in R&D efficiency within R&D distribution and try innovative research methods.

This article is based on endogenous development theory and mathematical statistics. When estimating total factor productivity, this article uses the Cobb-Douglas production function and Malmquist index methods to cover the weaknesses of a single model in the calculation process. The overall growth rate of factors calculated by these methods is basically the same.

This paper separates the fiscal market from the financial market. When analyzing the impact of the resource allocation efficiency of the tax market on total factor productivity, it combines theoretical research with practical research, enriching the current domestic and foreign scholars' research on the impact of fiscal R&D resource allocation efficiency on total factor productivity and theoretical research on the mechanism of factor productivity.

## 2. Impact of Total Factor Productivity under Complex Systems

**2.1. Complex System.** The definition of a complex system first appeared in the American journal "Science" magazine "Complexity Album." The two authors gave the following definition: by understanding the system's subsystems, it is impossible to fully explain the nature of the system. In other words, for a complex system, its total performance is not a simple superposition of partial performance, and there is a complex nonlinear relationship between total performance and partial performance [8].

Complex systems are the main research object of complexity science, and the synchronization phenomenon of complex systems is the current research focus. When studying complex systems, complex networks are considered to be important mathematical models, and their related theories are of great significance to the study of complex systems. Durability is a key attribute of dynamic systems and is usually used to describe the dynamic behavior of spaced systems. A retrospective graph is a visualization tool based on robustness. Quantitative robustness analysis is a measure of quantitative robustness analysis [9]. As a new method of studying complex systems, retrospective graphs and quantitative reduction analysis have been widely used in many complex system problems. The logical relationship between the theories used in this article is shown in Figures 1–3.

The concept of a complex system is realized based on the critical inheritance of reductionism. The theory of complex systems is not just a refusal to restore. Reductionism is valid and logical within a certain range, because in addition to complex systems, there are other types of systems in the world. Generally, there are three types of systems in the world: simple systems, random systems, and complex systems [10].

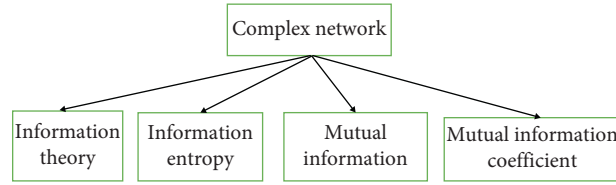


FIGURE 1: Complex system framework.

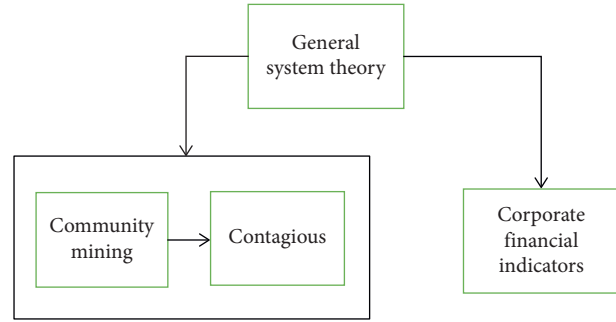


FIGURE 2: General system theory construction.

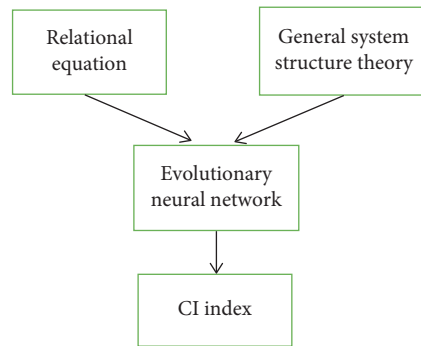


FIGURE 3: Logical relationship diagram of each theoretical tool.

A complex network is a mathematical model of a complex system. Its main purpose is to explain the existing phenomena and complexity of the network. The research object is the network [11]. The Internet is ubiquitous in nature and human society and has a profound impact on human development. Graph theory is an important tool for studying complex networks. Generally, a network is considered a graph, which contains a series of nodes and the ends that connect these nodes.

**2.2. VR Technology and Machine Learning.** VR technology is a virtual reality technology that uses computers to simulate real and unreal 3D scenes. Virtual reality technology is the highest level of current simulation and virtual reality, and it has integrated a variety of technologies, such as multimedia technology, digital image processing, sensor technology, and computer graphics [12]. VR technology can create 3D visual and sound effects. Since human-computer interaction is a friendly and harmonious state based on physical skills, 3D virtual reality technology can change the passive and boring state between man and machine. Virtual reality technology

combines computer technology, multimedia technology, image technology, simulation technology, and various electronic technologies to create new technology in the computer field, as shown in Figure 4.

Virtual reality technology is a computer simulation system that can build and experience the virtual world. Virtual reality technology is a very challenging interactive technology, which is gradually being widely used in the field of scientific research [13, 14]. At present, in the fields of military, medicine, film, and television, virtual reality technology has gradually become an important technology, and the development of virtual reality technology in the field of education is gradually deepening. Virtual reality technology is a cutting-edge technology in the computer field, which combines human-computer interaction technology, sensor technology, network technology, graphics technology, simulation technology, and stereo display technology. Virtual reality technology has three main characteristics, as shown in Figure 5.

Machine learning enables computers to learn and deal with human-like problems. Therefore, it is very important to focus on how to apply machine learning algorithms to solve

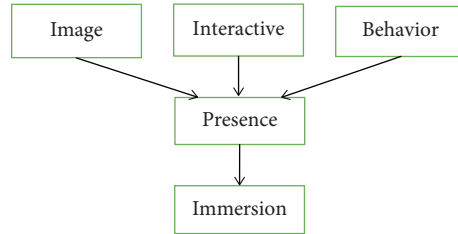


FIGURE 4: VR technology elements and characteristics.

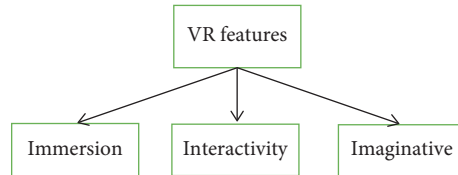


FIGURE 5: Three characteristics of virtual reality.

practical problems. Machine learning has a wide range of research scenarios and applications. It is a multidisciplinary field, including artificial intelligence, information theory, statistical probability, government, and other academic achievements [15, 16]. The machine learning algorithms involved in this research are Bayesian statistical conclusions, EM algorithms, and ADMM algorithms, as shown in Figure 6.

According to historical data, with the help of the dynamic grouping method, the indirect correlation between parameters is determined through the EM algorithm, which aims to realize the dynamic vector grouping support algorithm, which is beneficial to reduce the risk of information decision-making [17]. The specific technology application path is shown in Figure 7.

Machine learning uses detailed methods in statistics, evidence theory, neural networks, fuzzy sets, primitive sets, evolutionary computing, and other fields to complete data summarization, concept description, classification rule extraction, grouping analysis, data analysis, analysis and sequence pattern discovery, and other types of work [18, 19]. Compared with data mining, machine learning tends to provide different algorithms, while data mining tends to use these algorithms to solve problems. Common machine learning algorithms include the following categories:

- (1) Classification algorithm: representative algorithms are decision tree algorithm and a naive Bayes algorithm
- (2) Clustering algorithm: representative algorithms are K-means algorithm and EM algorithm
- (3) Association rule extraction algorithm: the representative algorithm is the Apriori algorithm
- (4) Support vector machine, namely, SVM algorithm
- (5) Neural Networks
- (6) Genetic algorithm

Logistic regression (LR) is a machine learning algorithm with low model complexity and a classic statistical

classification algorithm. Its advantages are low computational complexity, easy to understand and apply, and can be used for large-scale machine learning and online learning tasks [17]. The conditional probability distribution formula of the binomial logistic regression model is

$$P(A = 1|x) = \frac{\exp(mx + n)}{1 + \exp(mx + n)}, \quad (1)$$

$$P(A = 0|x) = \frac{1}{1 + \exp(mx + n)}.$$

*2.3. Fiscal R&D Resource Allocation Efficiency.* R&D (Research and Development) refers to the human, material, and financial resources required to participate in scientific research and experimental development activities. R&D expenditure and its proportion in GDP reflect the important content of our country's independent innovation capability and the process of building an innovative country. It is an important indicator to measure the scale of a country's scientific and technological activities and the level of scientific and technological investment, and it is also an important indicator to improve our country's innovation potential and enhance the competitiveness of the national economy. UNESCO believes that R&D is a systematic creative activity that uses new knowledge to create new applications. OECD (Organization for Economic Cooperation and Development) believes that R&D is a more systematic and creative activity. This type of activity is based on increasing the total amount of knowledge and using that knowledge to create new applications [20].

Resource allocation was first proposed by the classical economy. It emphasized the role of the market in the allocation of resources and believed that the market was the most important way to allocate resources. However, in fact, due to the shortcomings of the market itself, especially the failed market, it seriously wasted resources. Adam Smith believes that due to a lack of resources, the market uses

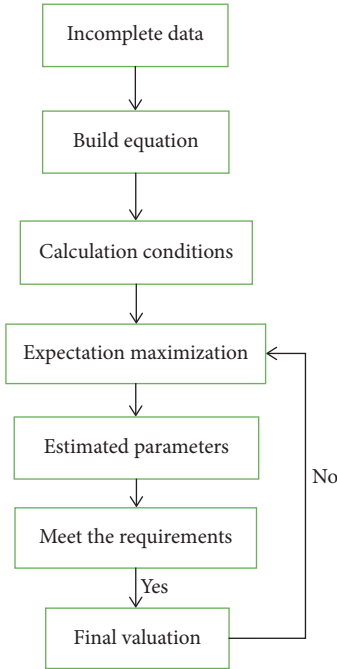


FIGURE 6: Flow chart of the EM algorithm.

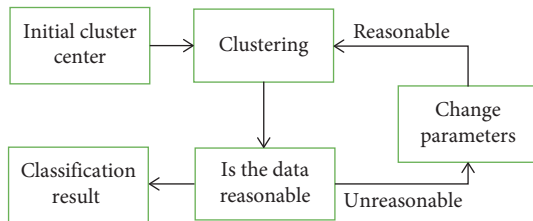


FIGURE 7: Technical route of dynamic clustering decision.

profit-driven methods to achieve effective resource allocation. Tables 1 and 2 list the average fiscal R&D resource allocation efficiency of countries and regions.

**2.4. TFP.** Total factor productivity (TFP) refers to factors other than capital and labor income that contribute to economic growth. It is an important indicator to measure the quality of economic growth. Our country has a lot of research on total factor productivity. Total factor productivity is proposed relative to the concept of single-factor productivity. After excluding the increase in input rate, the remaining contribution of production improvement sources is attributed to the improvement of technology and efficiency. This part is total factor productivity, that is, the rest of the economic growth that cannot be explained by input factors. Fundamentally speaking, total factor productivity measures the influence of factors such as technological innovation, process improvement, management level, and efficiency. The role of industrial organization in economic development reflects the quality of economic development and represents the overall efficiency of the use of input factors [21].

Comparing the differences in total factor productivity between different countries, the scale of total factor productivity will directly affect the level of economic growth. There are several methods for calculating TFP in empirical research. Solow residual algorithm will be used, and its approximate algorithm is as follows:

$$\text{TFP growth} = \frac{\dot{L}}{L} = \frac{\dot{M}}{M} - \beta \frac{\dot{N}}{N} - (1 - \beta) \frac{\dot{O}}{O}, \quad (2)$$

where  $L$  is the technology input,  $M$  is the output,  $N$  is the capital input,  $O$  is the labor input, and  $\beta$  is the proportion of capital input to output. According to the above Solow formula, if we want to be able to accurately calculate the growth of TFP, we must adopt the function of neoclassical production. The factor market is fiercely competitive, and the growth rate of various inputs must be accurate. In addition, many papers have adopted the following methods, but the basic idea is the same as the above formula:

$$\text{TFP} = \frac{M}{N^\beta \times O^{1-\beta}}. \quad (3)$$

In order to calculate total factor productivity, some documents divide the calculation methods into two categories: trend estimation method and production function method. Trend estimation methods include parametric, semiparametric, and nonparametric methods. It uses total factor productivity as the default variable and uses long-term statistical information to estimate total factor productivity based on a set of assumptions. The production function method is mainly based on the basic form of the Cobb-Douglas production function in the neoclassical development theory. It uses mathematical transformations to discover the number of factors affecting economic growth, such as technological level and technical performance, in addition to capital and labor. Therefore, the productivity of all factors can also be called the residual value of the production function [22].

This article will choose the SFA method and use the prior logarithmic function to calculate our country's interprovincial total factor productivity. The production function has the following form:

$$\begin{aligned} \ln S_{a,b} = & \alpha_0 + \alpha_1 \ln M_{ab} + \alpha_2 \ln N_{ab} + \beta_1 b \\ & + \frac{1}{2} \beta_2 \ln^2 N_{ab} + \frac{1}{2} \beta_3 \ln^2 M_{ab} \\ & + \frac{1}{2} \beta_4 b^2 + \beta_5 \ln N_{ab} M_{ab} + \beta_6 b \ln M_{ab} \\ & + \beta_7 b \ln N_{ab} + x_{a,b} - y_{a,b}. \end{aligned} \quad (4)$$

When all  $\beta$  is 0, the model will become a technically neutral Cobb-Douglas production function. In addition, the model also considers static and dynamic technological

TABLE 1: Average allocation efficiency of our country's fiscal R&amp;D resources.

| Year          | Technical efficiency | Technological progress rate | Pure technical efficiency | Scale efficiency | Allocation efficiency |
|---------------|----------------------|-----------------------------|---------------------------|------------------|-----------------------|
| 2014          | 1.041                | 0.967                       | 1.021                     | 1.018            | 0.999                 |
| 2015          | 0.988                | 1.030                       | 0.977                     | 1.021            | 1.003                 |
| 2016          | 1.032                | 0.944                       | 1.016                     | 1.017            | 0.956                 |
| 2017          | 0.999                | 0.947                       | 0.966                     | 1.052            | 0.955                 |
| 2018          | 1.011                | 1.069                       | 0.996                     | 1.032            | 1.095                 |
| 2019          | 1.036                | 0.981                       | 1.021                     | 1.025            | 1.008                 |
| Average value | 1.018                | 0.990                       | 1.000                     | 1.028            | 1.010                 |

TABLE 2: Average allocation efficiency of financial R&amp;D resources in each region.

| Year    | Technical efficiency | Technological progress rate | Pure technical efficiency | Scale efficiency | Allocation efficiency |
|---------|----------------------|-----------------------------|---------------------------|------------------|-----------------------|
| East    | 0.996                | 1.016                       | 0.996                     | 1.001            | 1.015                 |
| Central | 1.003                | 1.000                       | 0.998                     | 1.005            | 1.001                 |
| West    | 1.020                | 0.986                       | 1.018                     | 1.002            | 1.007                 |

progress and the dynamic interaction of labor and capital input into production. By organization, we get

$$IHK_{ab} = IJ_{ab} + IK_{ab} + (JF - 1)(\lambda_N(\lambda_N + \lambda_M)N + \lambda_M(\lambda_N + \lambda_M)M), \quad (5)$$

and decomposing it, the production efficiency change rate is

$$IJ_a = J[\exp(-y_a)|x_a - y_a]. \quad (6)$$

The technical progress rate is

$$IK_{ab} = \frac{\partial \ln S_{ab}}{\partial} = \beta_1 + \beta_4 b + \beta_6 \ln m_{ab} + \beta_7 \ln n_{ab}. \quad (7)$$

The rate of change of scale efficiency is

$$FJC = (JF - 1) \left( \frac{\mu_N}{\mu_N + \mu_M} \times N + (\mu_N + \mu_M)M \right). \quad (8)$$

Among them,  $\mu_N$  and  $\mu_M$ , respectively, represent the output elasticity of capital and labor relative to the overall return to scale:

$$\begin{aligned} \mu_N &= \alpha_N + \beta_7 b + \beta_5 \ln M_{ab} + \beta_2 \ln N_{ab}, \\ \mu_M &= \alpha_M + \beta_6 b + \beta_5 \ln N_{ab} + \beta_3 \ln M_{ab}. \end{aligned} \quad (9)$$

**2.5. Model Building.** In this study, the DEA-Malmquist index model was used to measure the effectiveness of R&D allocation in our country. In the DEA empirical research, in order to reflect the evolution of production unit efficiency in different periods, the most widely used production efficiency index is the Malmquist index [23]. Caves proposed that under multiple inputs and outputs, the total productivity based on the input data can be expressed by the Malmquist index. In order to obtain the Malmquist efficiency value, the distance function  $S^a(x_0^a, y_0^a)$  needs to be input, which is the inverse of the input efficiency, namely,

$$S^a(x_0^a, y_0^a) = \frac{1}{T^a(y_0^a, x_0^a)}. \quad (10)$$

This function is the smallest degree where  $x_0^a$  can be reduced under a given  $y_0^a$ . When

$$S^a(x_0^a, y_0^a) = 1, \quad (11)$$

$(x_0^a, y_0^a)$  on the frontier of production,

$$S^a(x_0^a, y_0^a) > 1 \quad (12)$$

indicates that the technology is invalid, and the performance value can be expressed by the Malmquist index:

$$T^a = \frac{S^a(x_0^a, y_0^a)}{S^a(x_0^{a+1}, y_0^{a+1})}, \quad (13)$$

$$T^{a+1} = \frac{S^{a+1}(x_0^a, y_0^a)}{S^{a+1}(x_0^{a+1}, y_0^{a+1})}.$$

These two indicators are the Malmquist index in the period  $a$  and period  $a + 1$ , respectively. We use the geometric mean of these two Malmquist indices to calculate the change in productivity, namely,

$$\begin{aligned} TI &= \frac{S^a(x_0^a, y_0^a)}{S^{a+1}(x_0^{a+1}, y_0^{a+1})} \left[ \frac{S^{a+1}(x_0^{a+1}, y_0^{a+1})}{S^a(x_0^{a+1}, y_0^{a+1})} \times \frac{S^{a+1}(x_0^a, y_0^a)}{S^a(x_0^a, y_0^a)} \right]^{1/2} \\ &= EF \times AC. \end{aligned} \quad (14)$$

$TI > 1$  shows that the productivity of the period  $t + 1$  is higher than the productivity of period  $a$ . Malmquist index can be decomposed into technical performance index and technological progress rate index. EF and AC are greater than, equal to, and less than 1, respectively, indicating that the technical performance and technological progress rate remain unchanged or decline, and the technical efficiency is

decomposed into pure technical efficiency index and index scale efficiency. The survey is used to measure the changes in the effectiveness of R&D allocation in 28 provinces and regions in our country from 2014 to 2019 [24].

A general method of estimating capital stock and current capital is the sum of current new capital investment and previous capital minus capital depreciation. The basic types of calculations are

$$R_x = \frac{A_x}{B_x} + (1 - \mu)R_{x-1}. \quad (15)$$

Among them,  $R$ ,  $A$ , and  $B$  are the capital stock, investment, and fixed asset investment price index,  $\mu$  is the annual depreciation rate of capital, and the subscript  $x$  represents the period [25].

### 3. Experiment on the Influence of Total Factor Productivity under Complex Systems

A large number of research results show that the efficiency of fiscal R&D resource allocation is a key factor in promoting overall productivity and economic growth. Comparing the results of different empirical methods, we can conclude that the massive influx of R&D funds based on international trade has greatly increased the overall factor productivity of our country's industry, while the low efficiency of domestic fiscal R&D resource allocation has restricted its growth. Therefore, it is recommended to improve the efficiency of fiscal R&D resource allocation and strengthen trade with developed countries.

**3.1. Test Subject.** The empirical part of this article first examines the direct impact of R&D resource allocation efficiency on improving the overall productivity of participants. In addition, in order to further determine the impact mechanism, this article decomposes the overall productivity of agents into changes in technical progress (TE) and technical efficiency (EF). Conversely, as explained by the variables, it empirically examines the impact of R&D resource allocation efficiency on technological progress and technical efficiency. Therefore, this paper constructs the following model with total factor productivity (TFP) as a variable explanation:

$$\begin{aligned} TFP_{ab} = & \omega + \lambda_1 TFP_{ab-1} + \lambda_2 FA_{ab-1} + \lambda_3 RND_{ab} + \lambda_4 HC_{ab} \\ & + \lambda_5 GI_{ab} + \lambda_6 FDI_{ab} + \lambda_7 OPEN_{ab} + \lambda_8 DAR_{ab} + z_{ab}. \end{aligned} \quad (16)$$

**3.2. Test Design.** Total factor productivity is the main indicator to measure the impact of resource allocation efficiency and technological innovation level on economic growth. In order to be able to estimate total factor productivity, we introduced the Cobb-Douglas (CD) production function into the analysis process based on the practice of other scholars. The production function is as follows:

$$S_{ab} = L_{ab} M_{ab}^\alpha N_{ab}^\beta, \quad (17)$$

where  $S_{ab}$  represents the GDP of province  $a$  in year  $b$ ,  $L_{ab}$  represents the total factor productivity of province  $a$  in year  $b$ , and  $M_{ab} N_{ab}$  represents the capital stock and labor force of province  $a$  in year  $b$ .  $\alpha$  and  $\beta$ , respectively, represent the output elasticity of the two production factors of capital and labor. Taking the logarithm of both sides of the formula equation at the same time, you can get

$$\ln S_{ab} = \ln L_{ab} + \ln M_{ab} + \ln N_{ab}. \quad (18)$$

Doing further transformation, you can get

$$\ln L_{ab} = \ln S_{ab} - \ln M_{ab} - \ln N_{ab}, \quad (19)$$

where  $\ln L_{ab}$  is the logarithm of the total factor productivity of each province over the years. Taking the physical indicators on the left and right sides of the equation at the same time, we can get the final expression of total factor productivity:

$$\ln L_{ab} = \exp(\ln S_{ab} - \ln M_{ab} - \ln N_{ab}). \quad (20)$$

We use the Solow residual method to decompose and measure total factor productivity and set the total production function as the C-D production function:

$$G_a = X e^{\lambda a} Y_a^\alpha Z_a^\beta. \quad (21)$$

Among them,  $G_a$  is the actual output,  $Z_a$  is the labor input,  $Y_a$  is the capital stock, and  $\alpha, \beta$  are the average labor output share and the average capital output share, respectively.

Now, we make the following assumptions: constant returns to scale; if the input quantity increases by the same percentage, the output also increases by the same percentage; the neutral technology assumes that labor productivity and capital productivity increase by the same percentage; the natural logarithm of the two sides is as follows:

$$\ln(G_a) = \ln(X) + \lambda a + \alpha \ln(Y_a) + \beta \ln(Z_a). \quad (22)$$

The growth rate of total factor productivity is derived as

$$\frac{\Delta X}{X} = \frac{\Delta G}{G - (1 - \alpha)} \times \frac{\Delta Z}{Z} - \alpha \times \frac{\Delta Y}{Y}. \quad (23)$$

Under the constraint condition  $\alpha + \beta = 1$  of constant return to scale, we have

$$\ln\left(\frac{G_a}{Z_a}\right) = \ln(X) + \lambda a + \alpha \ln\left(\frac{Y_a}{Z_a}\right). \quad (24)$$

After estimating the average share of capital production and labor productivity, the growth rate of total factor productivity can be obtained by introducing equations. We judge  $\alpha = 0.7, \beta = 0.3$  according to the empirical method and bring the total factor productivity table.

## 4. Impact on Total Factor Productivity

*4.1. Time Series Changes in the Efficiency of Fiscal R&D Resource Allocation.* As shown in Figure 8, the TFP growth rate of the entire industry from 2014 to 2019 was 8.4%, the average growth rate of technical performance was  $-1.4\%$ , and the average growth rate of technological progress was  $9.9\%$ . Technological progress plays a leading role in changes in total factor productivity. The contribution of technological progress to the overall productivity growth of the industry is far greater than changes in technical performance. Further analysis of the net changes in technical performance and scale performance changes of EF decomposition elements shows that my country as a whole has reduced the return to scale ( $-2.4\%$ ). The speed of technological progress is a key factor that affects the efficiency of R&D resource allocation. On this basis, the factor analysis method is used to analyze the environmental factors that affect the efficiency of R&D resource allocation. The research results show the government support, economic development level, and foreign investment level. It plays an important role in promoting my country's total factor productivity and is also an important reason for regional differences, as shown in Table 3.

As shown in Figure 9, the change in the efficiency of my country's fiscal R&D resource allocation is almost the same as the change in the pace of technological progress, while pure technical efficiency and scale performance have little effect on the efficiency of R&D allocation, indicating that technological progress at this pace directly affects the allocation of R&D resources. The overall efficiency of efficiency, the level of research, and development need to be improved. The efficiency of resource allocation must be based on the perspective of technological progress. As shown in Table 4, the rapid improvement of technological level, the renewal of technical equipment, technological innovation, and the introduction of advanced technology have a greater impact on improving the efficiency of my country's R&D resource allocation.

As shown in Figure 10, from 2014 to 2019, my country's overall resource allocation for fiscal research and development showed a slow growth trend, with an average rate of return of 1.01, but the growth rate was very slow. During these 7 years, the average growth rate of my country's R&D resource allocation efficiency was only  $0.5\%$ , which was mainly attributed to the slight increase in the rate of technological progress during the research period. In addition, the resource allocation efficiency of R&D reached the highest level in 2018, mainly due to the greatly improved technical level and the acceleration of technological progress. As shown in the figure, the slowing rate of technological progress has led to a decrease in the efficiency of R&D distribution, which is different in 2015 and 2018. The growth rate in 2016 has slightly decreased, as shown in Table 5.

*4.2. Regional Differences in the Efficiency of Fiscal R&D Resource Allocation.* As shown in Figure 11, in order to facilitate the analysis of regional differences in the efficiency of

my country's fiscal R&D resource allocation, this study divides my country's 30 provinces into eastern, central, and western regions. The R&D resource cost-sharing rate and national change coefficient of 30 provinces in eastern, central, and western regions of my country from 2000 to 2012 are given. The distribution efficiency of R&D resources in the eastern and central regions has shown a downward trend, while the western region has shown an upward trend.

As shown in Table 6, from the overall level, the efficiency of R&D resource allocation in each region is characterized by the highest in the east, the second in the west, and the lowest in the middle. The average efficiency is 1.014, 1.004, and 1.000 in order. The R&D distribution efficiency in the eastern region is 1.005 higher than the national average, and the central and western regions are lower than the national average. The coefficient of variation of the national average rate of return on R&D resources shows a downward trend, indicating that the regional differences in the efficiency of R&D resource allocation in my country are generally shrinking, as shown in Figure 12.

*4.3. Overall Situation of Total Factor Productivity.* As shown in Figure 13, total factor productivity decreased significantly from 2000 to 2016, and total factor productivity increased significantly from 2013 to 2014. This fully shows that my country's deepening reform and opening-up have led to economic growth and total factor productivity. Total factor productivity reached its peak in almost every city in 2014 and then declined every year. This shows that reform and opening-up can indeed bring technological progress and increase overall factor productivity, but with the past of reform and opening-up, this impact will continue to weaken, as shown in Table 7.

As shown in Figure 14, from a national level, the overall factor productivity generally declined. From 2010 to 2019, the growth rate of the total factor productivity in most cities was less than zero in most years. This shows that my country's economic growth is highly dependent on capital and labor, and my country's technological progress is not obvious. On average, my country's major cities have shrunk by  $5.78\%$ . During these years, the total factor productivity of 41% of the cities increased in 2010, which was the best year for total factor productivity. However, the overall factor productivity in 2015, 2018, and 2019 showed an overall downward trend. Almost 95% of cities showed negative growth in all factors.

*4.4. TFP and Decomposition Index.* As shown in Figure 15, using DEAP2.1 software, we select the input-oriented fixed income statement model and use table data from 30 provinces to obtain the results of the Malmquist index method. The Malmquist index decomposition model provides a comprehensive performance index for the total factor productivity of each province in my country. The index can be decomposed into the rate of scale efficiency change and the net change of technical performance according to the rate of change of technical performance, the rate of technological progress, and the rate of change of technical

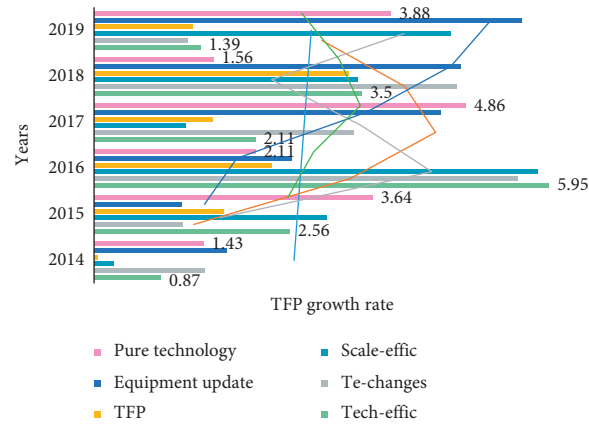


FIGURE 8: Net changes in technical performance and changes in scale performance.

TABLE 3: 2014–2019 total factor productivity index.

| Year    | Technical efficiency changes | Technological progress changes | Scale efficiency changes | TFP   |
|---------|------------------------------|--------------------------------|--------------------------|-------|
| 2014    | 1.101                        | 0.987                          | 1.070                    | 1.086 |
| 2015    | 0.925                        | 1.236                          | 0.945                    | 1.142 |
| 2016    | 0.886                        | 1.135                          | 0.905                    | 1.006 |
| 2017    | 1.014                        | 1.041                          | 1.001                    | 1.055 |
| 2018    | 0.988                        | 1.038                          | 0.999                    | 1.025 |
| 2019    | 1.005                        | 1.028                          | 1.002                    | 1.032 |
| Average | 0.987                        | 1.078                          | 0.987                    | 1.058 |

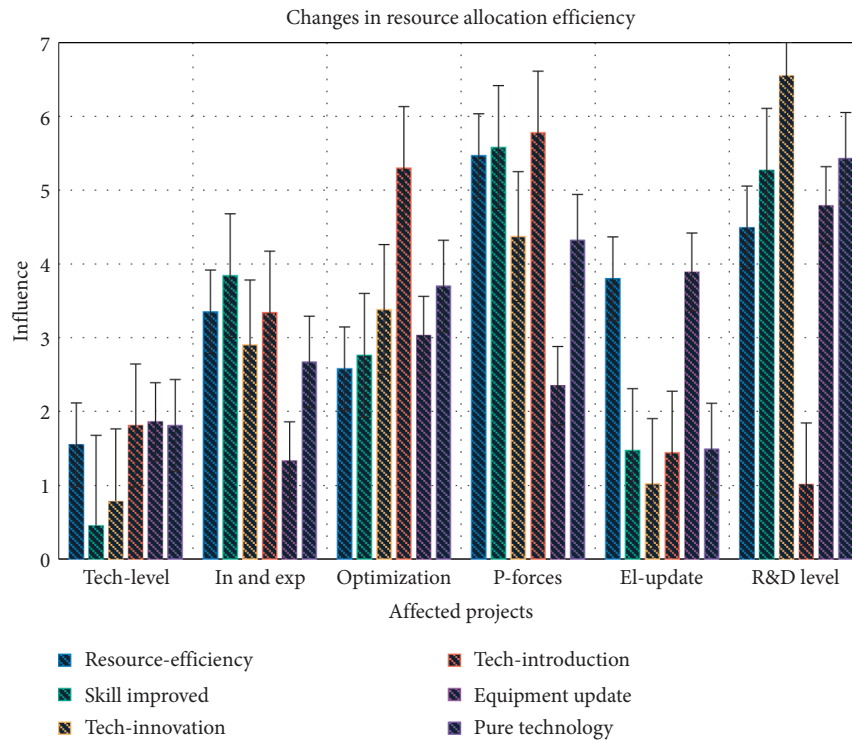


FIGURE 9: Relationship between changes in resource allocation efficiency.



TABLE 4: R&D resource allocation efficiency has a greater impact.

| Item         | Resource-efficiency | Innovation | Skill | Introduction | Equipment | Pure |
|--------------|---------------------|------------|-------|--------------|-----------|------|
| Tech-level   | 1.55                | 0.78       | 0.45  | 1.81         | 1.86      | 1.81 |
| In and exp   | 3.35                | 2.9        | 3.84  | 3.34         | 1.33      | 2.67 |
| Optimization | 2.58                | 3.38       | 2.76  | 5.3          | 3.03      | 3.7  |
| P-forces     | 5.47                | 4.37       | 5.58  | 5.78         | 2.35      | 4.32 |
| El-update    | 3.8                 | 1.02       | 1.47  | 1.44         | 3.89      | 1.49 |
| R&D level    | 4.49                | 6.55       | 5.27  | 1.01         | 4.79      | 5.43 |

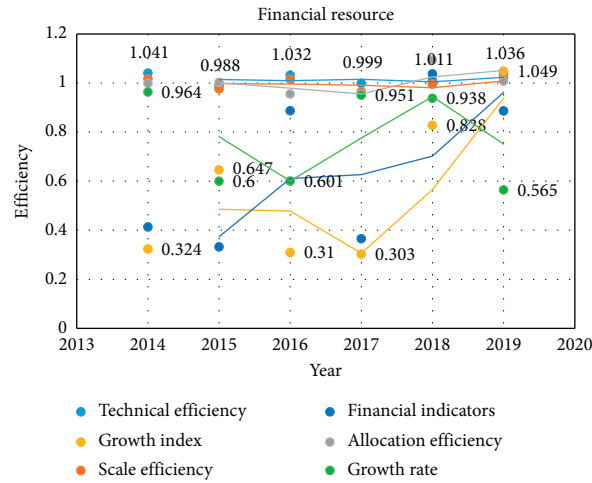


FIGURE 10: The efficiency of financial resource allocation in recent years.

TABLE 5: Decreased rate of technological progress leads to lower R&D distribution efficiency.

| Item | Technical efficiency | Scale efficiency | Allocation efficiency | Growth index | Financial indicators | Growth rate |
|------|----------------------|------------------|-----------------------|--------------|----------------------|-------------|
| 2014 | 1.041                | 1.018            | 0.999                 | 0.362        | 0.403                | 0.708       |
| 2015 | 0.988                | 0.977            | 1.003                 | 1.081        | 0.454                | 0.849       |
| 2016 | 1.032                | 1.016            | 0.956                 | 0.796        | 0.743                | 0.651       |
| 2017 | 0.999                | 0.966            | 0.955                 | 0.827        | 0.666                | 0.41        |
| 2018 | 1.011                | 0.996            | 1.095                 | 0.348        | 0.999                | 0.354       |
| 2019 | 1.036                | 1.021            | 1.008                 | 0.488        | 0.566                | 0.313       |

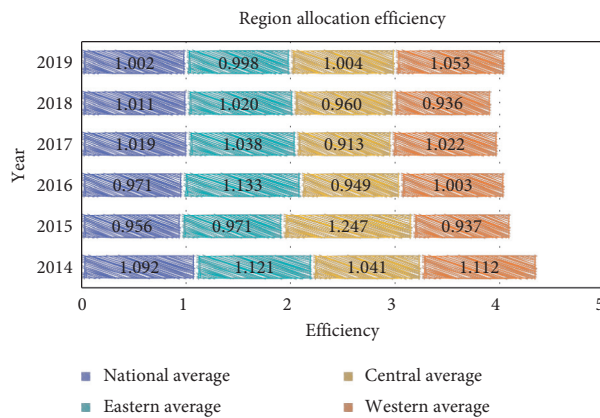


FIGURE 11: Changes in the efficiency of regional fiscal R&D resource allocation.

TABLE 6: R&D resource allocation efficiency in various regions.

|         | Distribution rate | Allocation efficiency | Response rate | Mean efficiency | R&D rate | Factors of production |
|---------|-------------------|-----------------------|---------------|-----------------|----------|-----------------------|
| East    | 1.84              | 0.34                  | 1.88          | 0.94            | 1.84     | 0.67                  |
| West    | 1.52              | 2.79                  | 3.84          | 3.19            | 3.57     | 2.7                   |
| Central | 2.96              | 4                     | 3.36          | 5.07            | 4.46     | 3.53                  |
| South   | 2.25              | 2.66                  | 2.74          | 4.66            | 1.51     | 4.75                  |
| North   | 2.98              | 1.64                  | 1.78          | 1.58            | 1.58     | 2.49                  |

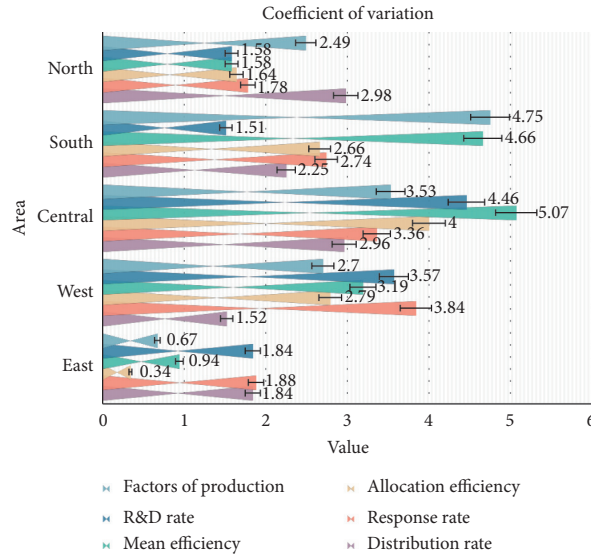


FIGURE 12: Coefficient of variation of the average rate of return on national R&D resources.

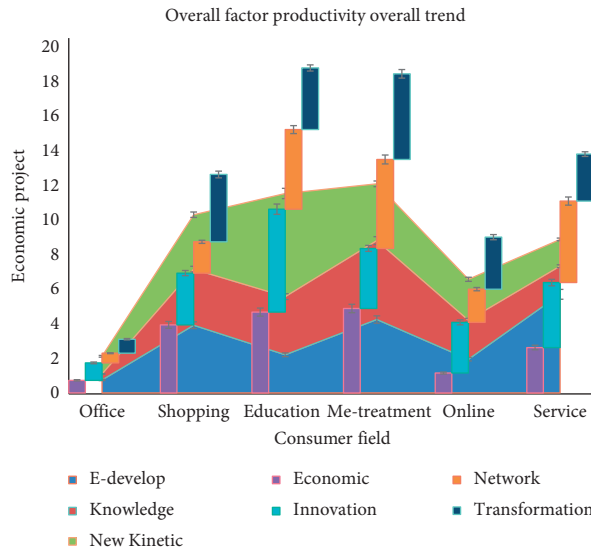


FIGURE 13: Overall factor productivity overall trend.

performance rate. As shown in Table 8, from 2014 to 2019, my country's total factor productivity increased by 4.02% annually, net efficiency changed (-0.37%), net technological progress changed (3.91%), scale efficiency changed (-0.08%), and technology scale changed (-0.37%).

Technological progress is the main factor leading to the increase in green total factor productivity. The regional differences in my country's green factor productivity growth are very obvious. From east to west, there is a gradual downward trend. From 2014 to 2019, the average annual

TABLE 7: The influence of total factor productivity.

| E-Item       | E-develop | New kinetic | Knowledge | Economic | Innovation | Network | Transformation |
|--------------|-----------|-------------|-----------|----------|------------|---------|----------------|
| Office       | 0.75      | 1.02        | 0.35      | 0.73     | 1.01       | 0.56    | 0.81           |
| Shopping     | 3.92      | 3.15        | 3.25      | 3.94     | 3          | 1.81    | 3.9            |
| Education    | 2.21      | 5.97        | 3.37      | 4.68     | 5.96       | 4.61    | 3.57           |
| Me-treatment | 4.25      | 3.29        | 4.57      | 4.89     | 3.48       | 5.15    | 4.96           |
| Online       | 1.95      | 2.37        | 2.26      | 1.15     | 2.94       | 1.92    | 3.01           |
| Service      | 5.71      | 1.55        | 1.63      | 2.63     | 3.76       | 4.72    | 2.72           |

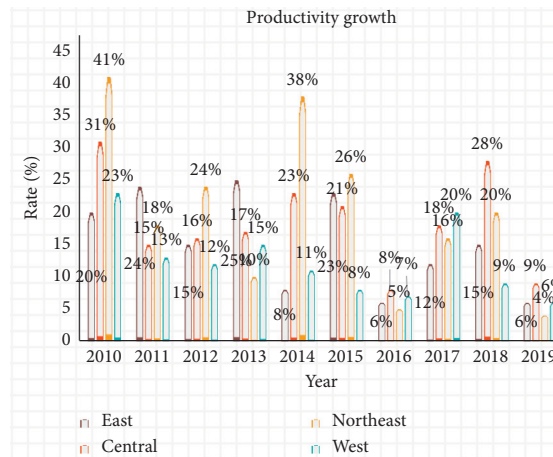


FIGURE 14: Proportion of cities with total factor productivity growth in each region.

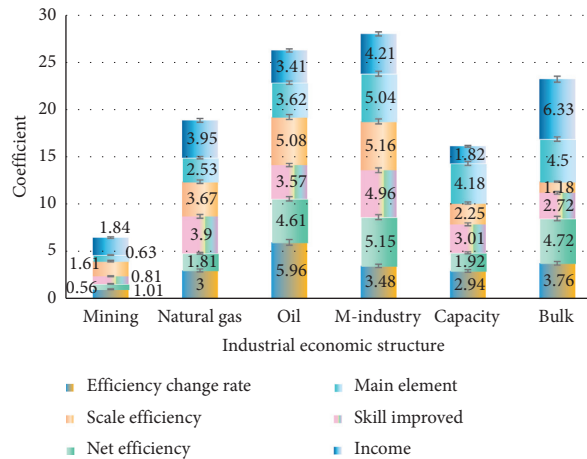


FIGURE 15: Results of the Malmquist index method.

growth rate of total factor productivity in the eastern region was 5.56%, the central region was 3.60%, and the western region was 2.79%, as shown in Table 9.

4.5. Calculation and Analysis of TFP in the East, Middle, and West. As shown in Figure 16, in order to avoid duplication of the work of this article when analyzing the TFP differences in the three main regions of eastern, central, and western regions, this article studies the differences in these three regions based on the results of the Malmquist index calculation method. In order to visually see the

changes in total factor productivity in the three main regions, the average total factor productivity value of each region and province is used. In the case of subregional levels, the overall growth rate of agents in the eastern region is the fastest. Secondly, the overall growth rate of agents in the central region is usually higher than that in the western region.

As shown in Figure 17, the development of total factor productivity in these three regions from 2004 to 2009 was mainly attributed to technological progress. After 2009, this is the result of the comprehensive influence of technological

TABLE 8: The index is based on the rate of change of technical performance.

| Item        | Efficiency change rate | Net efficiency | Skill improved | Scale efficiency | Main element |
|-------------|------------------------|----------------|----------------|------------------|--------------|
| Mining      | 1.01                   | 0.56           | 0.81           | 1.61             | 0.63         |
| Natural gas | 3                      | 1.81           | 3.9            | 3.67             | 2.53         |
| Oil         | 5.96                   | 4.61           | 3.57           | 5.08             | 3.62         |
| M-industry  | 3.48                   | 5.15           | 4.96           | 5.16             | 5.04         |
| Capacity    | 2.94                   | 1.92           | 3.01           | 2.25             | 4.18         |
| Bulk        | 3.76                   | 4.72           | 2.72           | 1.18             | 4.5          |

TABLE 9: Economic implications of TFP and its breakdown indicators.

| Index                       | Economic implications  | >0                               | <0                                      |
|-----------------------------|--|----------------------------------|---|
| TFP                         | Total factor productivity after deducting environmental costs                      | Total factor productivity growth | Green total factor productivity decline |
| Pure efficiency change      | Management innovation, system innovation<br>Induced productivity changes           | Pure efficiency improvement      | Pure efficiency deterioration           |
| Pure technological progress | Technological innovation, process improvement<br>Induced productivity changes      | Pure technological progress      | Pure technology regression              |
| Scale efficiency changes    | Productivity changes caused by economies of scale brought about by scale expansion | Increased scale efficiency       | Scale efficiency decline                |
| Technological scale changes | It only represents the technical meaning of DEA, not the economic meaning          | Technology deviation from CRS    | Technology deviation from VRS           |

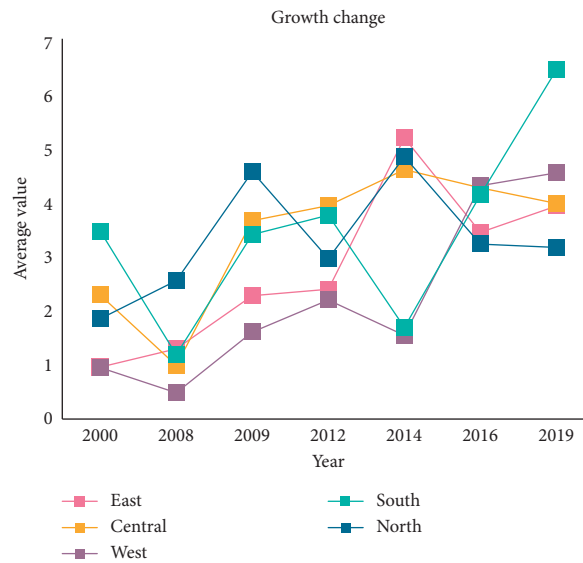


FIGURE 16: Average TFP of east, mid, and west.

progress and technical efficiency, making total factor productivity an annual trend. After 2000, the macroeconomy entered a stage of improvement. The technological progress in the eastern region has always been higher than that in the central and western regions, followed by the technological progress in the central region, and the lowest in the western region. This is the result of the spread of technological

progress. While taking advantage of the geographical advantages of the neighboring eastern region, the central region of my country must introduce advanced technologies, maintain technological progress, and at the same time pay attention to improving technical efficiency to more effectively promote the development of total factor productivity, as shown in Table 10.

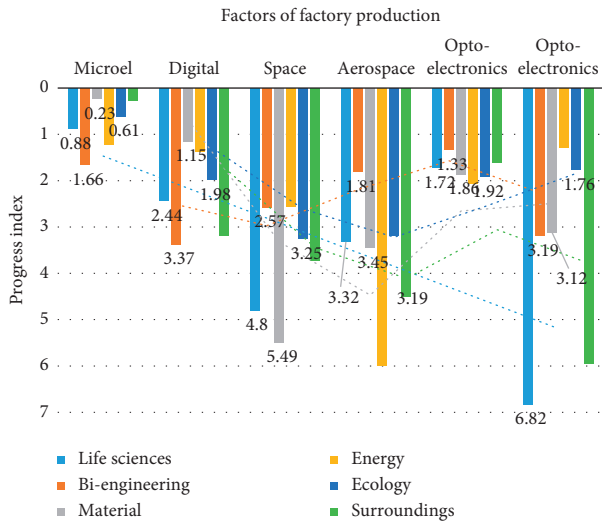


FIGURE 17: The macroeconomy has entered an improvement stage.

TABLE 10: Differences in TFP between the three main regions in the central and western regions.

| Years | East  | Central | West  | South | North |
|-------|-------|---------|-------|-------|-------|
| 2000  | 0.968 | 2.31    | 0.958 | 3.5   | 1.87  |
| 2008  | 1.31  | 1       | 0.49  | 1.2   | 2.58  |
| 2009  | 2.3   | 3.7     | 1.62  | 3.44  | 4.61  |
| 2012  | 2.42  | 3.98    | 2.22  | 3.8   | 2.99  |
| 2014  | 5.24  | 4.65    | 1.56  | 1.71  | 4.89  |
| 2016  | 3.48  | 4.31    | 4.35  | 4.18  | 3.26  |
| 2019  | 3.97  | 4.02    | 4.59  | 6.51  | 3.2   |

### 5. Conclusion

In the short term, the efficiency of fiscal R&D resource allocation promotes the development of total factor productivity, but as time goes by, its effect weakens. TFP limits the quality of economic growth in the short term, and the long-term effect of promoting the quality of economic growth tends to increase; during the inspection period, the economic growth rate is restricted. Among the three economic growth variables studied, the speed of economic growth has the greatest impact on the quality of economic growth. Therefore, at this stage, our country should control the rate of economic growth, improve the rationality of the allocation of R&D resources, promote the growth of total factor productivity, and achieve a slowdown in economic growth and high-quality growth.

Improving the quality of total factor productivity and economic growth will in turn help the rational allocation of fiscal R&D resources. The impact of economic growth on this has changed from positive to negative, which indicates that the excessive pursuit of economic growth will be ignored. This shows that excessive pursuit of economic growth will be ignored and to a certain extent will lead to waste of resources. The quality of economic growth has a deterrent effect on total factor productivity in the short term, but in the long run, it has an increasingly stronger role in promoting total factor productivity. Among these three variables, the

quality of economic growth has an effect on total factor productivity, which is the greatest contribution. Therefore, our country's pursuit of low-speed, high-quality economic growth can feedback the rational allocation of fiscal R&D resources and the development of total factor productivity, which forms a virtuous circle of economic activities.

If our country blindly pursues economic growth at the expense of the environment and resource waste, it will lead to a shortage of various social and natural resources, which will lead to a financial crisis, inflation, weak economic growth, and even serious problems. Excessive pursuit of the quality of economic growth requires increasing welfare benefits, increasing the cost of controlling environmental pollution, and shutting down a large number of seriously polluting enterprises. This poses a huge challenge to our country's fiscal expenditure capacity and leads to a slowdown in economic growth. Therefore, the relationship between the two must be properly handled to ensure healthy and stable economic growth.

### Data Availability

No data were used to support this study.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### References

- [1] G. Cerulli, B. Potì, and R. Spallone, "The impact of fiscal relief on multinationals business R&D investments: a cross-country analysis," *Economia Politica*, vol. 35, no. 2, pp. 649–675, 2018.
- [2] L. Zhao, M. Du, and L. Chen, "A new multi-resource allocation mechanism: a tradeoff between fairness and efficiency in cloud computing," *China Communications*, vol. 15, no. 3, pp. 57–77, 2018.
- [3] A. Dehnokhalaji, M. Ghiyasi, and P. Korhonen, "Resource allocation based on cost efficiency," *Journal of the Operational Research Society*, vol. 68, no. 10, pp. 1279–1289, 2017.
- [4] B. Jia, H. Hu, Y. Zeng, T. Xu, and Y. Yang, "Double-matching resource allocation strategy in fog computing networks based on cost efficiency," *Journal of Communications and Networks*, vol. 20, no. 3, pp. 237–246, 2018.
- [5] D. Furceri and J. T. Jalles, "Fiscal counter-cyclicality and productive investment: evidence from advanced economies," *BE Journal of Macroeconomics*, vol. 19, no. 1, pp. 1–15, 2019.
- [6] G. Liu, B. Wang, and N. Zhang, "A coin has two sides: which one is driving China's green TFP growth?" *Economic Systems*, vol. 40, no. 3, pp. 481–498, 2016.
- [7] C. T. Tugcu and A. K. Tiwari, "Does renewable and/or non-renewable energy consumption matter for total factor productivity (TFP) growth? Evidence from the BRICS," *Renewable and Sustainable Energy Reviews*, vol. 65, pp. 610–616, 2016.
- [8] C. Zheng, J. Le Duigou, M. Bricogne, and B. Eynard, "Design process for complex systems engineering based on interface model," *Insight*, vol. 18, no. 4, pp. 22–24, 2016.
- [9] B. S. Kim and T. G. Kim, "Cooperation of simulation and data model for performance analysis of complex systems," *International Journal of Simulation Modelling*, vol. 18, no. 4, pp. 608–619, 2019.

- [10] S. Nichele, *Introduction to the Modeling and Analysis of Complex Systems*, H. Sayama, Ed., p. 498, Open SUNY Textbooks, Binghamton, NY, USA, 2015.
- [11] Q. Bai and F. Ren, "Editorial special issue on agent-based modelling for complex systems," *Journal of Systems Science and Systems Engineering*, vol. 27, no. 2, pp. 130–133, 2018.
- [12] A. Menin, R. Torchelsen, and L. Nedel, "An analysis of VR technology used in immersive simulations with a serious game perspective," *IEEE Computer Graphics and Applications*, vol. 38, no. 2, pp. 57–73, 2018.
- [13] L. Sai and H. Yufei, "Study on the architecture design and interior decoration based on VR technology and computer simulation platform," *Paper Asia*, vol. 35, no. 2, pp. 54–57, 2019.
- [14] H. Hou and H. Wu, "Technology for real estate education and practice: a VR technology perspective," *Property Management*, vol. 38, no. 2, pp. 311–324, 2020.
- [15] S. Loftin, "Machine learning vs conventional analysis techniques for the earth's magnetic field study," *SMU Data Science Review*, vol. 2, no. 1, p. 7, 2019.
- [16] B. Zhang, X. Wan, F. S. Ouyang et al., "Machine learning algorithms for risk prediction of severe hand-foot-mouth disease in children," *Scientific Reports*, vol. 7, no. 1, pp. 53–68, 2017.
- [17] A. Radovic, M. Williams, D. Rousseau et al., "Machine learning at the energy and intensity frontiers of particle physics," *Nature*, vol. 560, no. 7716, pp. 41–48, 2018.
- [18] J. Jiang, L. Yu, J. Jiang et al., "Angel: a new large-scale machine learning system," *National Science Review*, vol. 5, no. 2, pp. 102–122, 2017.
- [19] E. Giacomidis, A. Matin, J. Wei, N. J. Doran, L. P. Barry, and X. Wang, "Blind nonlinearity equalization by machine-learning-based clustering for single- and multichannel coherent optical OFDM," *Journal of Lightwave Technology*, vol. 36, no. 3, pp. 721–727, 2018.
- [20] J. Guan, K. Zuo, K. Chen, and R. C. M. Yam, "Does country-level R&D efficiency benefit from the collaboration network structure?" *Research Policy*, vol. 45, no. 4, pp. 770–784, 2016.
- [21] Ş. Yazgan and Ö. Yalçinkaya, "The effects of research and development (R&D) investments on sustainable economic growth: evidence from OECD countries (1996–2015)," *Review of Economic Perspectives*, vol. 18, no. 1, pp. 3–23, 2018.
- [22] H. Xie, Q. Chen, F. Lu, W. Wang, G. Yao, and J. Yu, "Spatial-temporal disparities and influencing factors of total-factor green use efficiency of industrial land in China," *Journal of Cleaner Production*, vol. 207, pp. 1047–1058, 2019.
- [23] A. Shahabadi, F. Kimiaei, and M. Arbab Afzali, "The evaluation of impacts of knowledge-based economy factors on the improvement of total factor productivity (a comparative study of emerging and G7 economies)," *Journal of the Knowledge Economy*, vol. 9, no. 3, pp. 896–907, 2018.
- [24] C. Tsamadias, P. Pegkas, E. Mamatzakis et al., "Does R&D, human capital and FDI matter for TFP in OECD countries?" *Economics of Innovation and New Technology*, vol. 28, no. 3–4, pp. 386–406, 2019.
- [25] J.-H. Lin, S. Chen, S. Chen, and J.-Y. Tsai, "How does soft information about small business lending affect bank efficiency under capital regulation?" *Quantitative Finance and Economics*, vol. 3, no. 1, pp. 53–74, 2019.

## Research Article

# Simulation Study on Fire Visibility of Typical Floor Planes of Modern Super High-Rise Office Buildings in China

Tongtong Zhang 

*School of Architecture & Urban Planning, Shenzhen University, Nanshan District, Shenzhen 518060, China*

Correspondence should be addressed to Tongtong Zhang; zhangtt@szu.edu.cn

Received 16 September 2020; Revised 31 October 2020; Accepted 23 November 2020; Published 28 December 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Tongtong Zhang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of office forms, the space form of super high-rise office buildings changed from the unitary efficient office space to a complex space that integrated office, communication, and experience, which also diversified the design of typical floors in the office zone. However, from the perspective of fire prevention, the placement of shared space changed the form of the plane in typical floors in the office zone, affecting the smoke spreading of fire and paths of personnel evacuation. Hence, the subject on the planar relationship among high-rise office buildings based on fire prevention analysis, which optimizes space design, is worthy of discussion. After collecting many cases of super high-rise office buildings in China, this study categorized them into six typical planes and adopted the software PyroSim for comparative simulation of the smoke spreading of fire. By comparing the visibility of different zones in fires, this study analyzed the effects of the area of the office zone, the location of the atrium, and the form of the plane on the fire visibility on the typical floor and put forth the key factors that influence fire visibility, thus optimizing the plane design of the typical floor of super high-rise buildings. The findings show that in the six fire scenes, the area of the office zone is between  $1136\text{m}^2$  and  $1736\text{m}^2$ . The peak of duration for visibility decline at 1.5 m appeared in Scene 3. Its office zone is  $1536\text{m}^2$ , and the duration of visibility decline is greater than 1000 s. In other fire scenes, the duration of visibility decline is less than 300 s. By comparing the plane arrangement features of the fire scenes, the paper concludes that in a given fire scene, establishing an atrium in the office zone, expanding the length of the adjacent edge between the atrium and the office zone, and building an evacuation corridor between the atrium and the core tube can greatly extend the effective evacuation time. This can serve as reference for the plane design of super high-rise office buildings.

## 1. Introduction

Super high-rise office buildings shoulder the important mission of multilayered optimization and development. Advancements in structural engineering have arisen to make possible the increase in height size and complexity, the reduction of cost and carbon footprint, and architectural imagination and economic versatility of these buildings [1]. As the modern office model becomes increasingly integrated, open, and smart, the plane design of typical floors, on the one hand, has increased the utilization value of the building space. But on the other hand, it has caused new fire safety problems to the buildings and resulted in greater difficulty in the fire safety design of the typical floor of buildings (Torero et al. [2]).

From both functional and structural perspectives, the typical floor plane of super high-rise office buildings, in general, can be roughly divided into two parts—the service part featuring core tubes and the service-receiving part characterized by frameworks. Judging from the development trend of existing structural techniques and the area of the typical floor, the upper limit of the ideal area of a typical floor can be lifted to  $4000\text{m}^2$  in a predictable future. After analyzing and investigating into the plane design drawings of super high-rise office buildings in recent years, this study has found that the floor space of an open and integrated office model needs to be as large as possible and continuous. Maximizing the plane area of the typical floor and exploring open office compartments are the two trends in the plane design of these buildings in modern times. For an office

space, a greater pillar spacing provides a broader view. As a design principle followed by the buildings of this kind in modern times, the pillar-free design indicates that in a designated office space, all bearing pillars are located near the curtain walls on all sides of the space to create a pillar-free space to provide a more flexible and efficient office zone. The correct definition of the design fire for open plan compartments is identified as the critical knowledge gap that must be addressed in order to achieve tall building performance objectives and to provide truly innovative, robust fire safety for these unique structures [1].

We have seen how classic prescriptive solutions failed to manage smoke (Cook Country Building (USA) [3] and Camberwell fire (UK) [4] and how modern buildings using state-of-the-art fire engineering failed to contain the full propagation of a fire (TVCC, China) [5]. Analyses of several failures and current design practices reveal that fire safety codes are no longer capable of providing implicit safety for the rapidly evolving needs of modern tall buildings and are being extensively substituted by nonvalidated performance-based design methods [1, 6]. Therefore, how to maximize the office zone on the typical floor and the space utilization efficiency while meeting the technical requirements on fire safety is a difficulty in typical floor design (Xing et al. [7]).

### 1.1. Fire Load of Super High-Rise Office Buildings.

Usually, massive office equipment, materials, papers, and archives are placed in an office. All these are highly combustible objects, thus imposing a heavy fire load on the office zone. Fire load ( $Q_f$ ) refers to both the heat of combustion of all combustible objects in the office zone and within the regional space. It is directly proportional to fire risks and the difficulty of firefighting (Yao et al. [8]). The total fire load of a unit area is fire load density ( $q$ ), which shows the severity of fire. Table 1 demonstrates the statistics on the average levels of fire load density in various functional space types of accredited buildings.

The computational analysis of the fire growth model effectively controlled smoke spread and emission.  $t^2$  model, MRFC model, and FFB model are the mathematical models of fire growth [10]. According to the results of numerous experiments and experience, the  $t^2$  model is used to describe the early stages of fire development in a specific space. Equation (1) is according to NFPA204M (2002):

$$Q_f = \alpha t^2, \quad (1)$$

where  $Q_f$  is the heat release rate(HRR), kW, and  $\alpha$  is the fire increasing modulus, kW/s<sup>2</sup>.

Table 2 shows the  $t^2$  model of different fire growth coefficients ( $\alpha$ ) adopted by various building designs and the maximum heat release rate and time of different materials.

### 1.2. Smoke Discharge of Office Zone in Super High-Rise Buildings.

According to the plane features and development trend of super high-rise office buildings, the demand for a pillar-free, open space has exceeded the maximum for a fire safety zone as required in the standards, which makes it

TABLE 1: Fire load density of different types of buildings [9].

|        | Space function | Density of combustible material |
|--------|----------------|---------------------------------|
| Office | General        | 30                              |
|        | Design         | 50                              |
|        | Administration | 60                              |
|        | Research       | 60                              |

impossible to achieve the fire safety zoning according to the traditional fire safety design. A higher office zone has a higher wind speed which will accelerate fire spreading on the typical floor in case of fire. Worse still, the absence of division in the interior lateral space would speed up the spreading of smoke [11].

According to the Code for Fire Protection Design of Building (GB50016-2014), "Mechanical smoke exhaust facilities should be installed and sites with interior net story height less than 6 m should include smoke control zones; the area of structure in each smoke control zone should not exceed 500 m<sup>2</sup>; and the smoke control zone should not cross beyond the fire protection zone. The smoke control zone should be separated by structural beams of partitions and ceilings with downward bulges no less than 500 m, or incombustible components of ceilings and canopies with downward bulges no less than 500 m." The smoke exhaust volume of mechanical smoke exhaust devices is demonstrated in Table 3.

Hence, it can be concluded that the area ( $S$ ) of each outlet is obtained with the following equation:

$$S = \frac{S1 \times 60 \text{ m}^3/\text{h}}{n \times 3600 \times 5 \text{ m/s}}, \quad (2)$$

where  $S$  is the area of the smoke discharge outlet,  $S1$  is the area of the smoke safety zone, and  $n$  is the number of smoke discharge outlets.

The area ( $s$ ) of each outlet was calculated according to the volume of the atrium through the following equation:

$$S = \frac{V \times 6}{n \times 3600 \times 10 \text{ m/s}}, \quad (3)$$

where  $S$  the area of the smoke discharge outlet,  $V$  is the volume of the atrium, and  $n$  is the number of smoke discharge outlets.

### 1.3. Crowd Evacuation of Super High-Rise Buildings.

Crowd evacuation in super high-rise buildings is a major safety concern [12]. According to the standards (GB50016-2014), the typical floor must provide two evacuation routes in case of fire, so that those trapped in fire can take the other if either of the routes is blocked (Wang and Liu [13]). For a more spacious typical floor, evacuation routes are longer present, and the choice of routes is blurrier, which would affect evacuation (Rodrigo and Marshall [14]). A subsafety zone can be established in the atrium to alleviate the pressure of evacuation (Liu [15]). But, the fire danger of the atrium is special, so if the subsafety zone is established in an inappropriate way, it will not reduce the pressure of evacuation and will even accelerate the vertical spreading of smoke.



TABLE 2: Fire growth coefficient (NFPA204M(2002)).

| Fire categories | $\alpha$ (kW/s <sup>2</sup> ) | $Q_f$ = the time of 1000 kW/s | Functions   |
|-----------------|-------------------------------|-------------------------------|---|
| Slow fire       | 0.0029                        | 600                           | Art gallery   |
| Medium fire     | 0.012                         | 300                           | Dwelling, apartment, construction room,<br>hotel reception, hotel bedroom |
| Fast fire       | 0.047                         | 145                           | Store   |
| Superfast fire  | 0.187                         | 75                            | —   |

TABLE 3: Minimum smoke exhaust volume of the mechanical system.

| Conditions and parts   | Unit smoke exhaust volume (m <sup>3</sup> (h))    | Ventilation (time/h) | Notes   |
|--|---|----------------------|---|
| One section of smoke control zone  | 60  | —                    | The smoke exhaust volume of each draught fan should not be less than 7200 m <sup>3</sup> /h                     |
| Areas with no smoke control zone and interior net story height greater than 6m |   |                      |   |
| Two or more smoke control zones  | 120   | —                    | Determined according to the maximum area of smoke control zone  |
| Atrium   | Volume less than or equal to 17000 m <sup>3</sup> | —                    | Volume greater than 17000 m <sup>3</sup> , smoke exhaust volume should be no less than 102000 m <sup>3</sup> /h |
|  | Volume greater than 17000 m <sup>3</sup>          | —                    |   |
|  |   | 6                    |   |
|  |   | 4                    |   |

Compared with CO mass fraction and temperature, visibility was a more influential factor in determining the critical time required for fire to become a hazard, and smoke affected the adjacent open area in approximately 60 s [12]. When a fire hazard is occurring, office zone visibility can directly affect the velocity of walking. In the irritating state and nonirritating state, the velocity of walking declines as the obscuration coefficient increases. In the irritating state, the velocity of walking showed sudden decline. Hence, the visibility of the horizontal section (1.5 m of interface) at the height of sight of crowd evacuation is one of the effective evaluation criteria about whether a certain area fits the condition of evacuation. In simulated calculations, the critical time of a dangerous situation is time  $T_4$  for visibility to decline to 10 m [16].

## 2. Materials and Methods

This research uses the software PyroSim to simulate the smoke spreading of fire [17]. The simulation process includes the following steps: (1) Building geometric models; (2) Establishing fire scenarios, including location and load of combustion source, simulation scope and boundary condition, different types of combustible materials, performance of firefighting equipment, duration and accuracy of simulation; and (3) conducting simulated calculation through FDS/smokeview for posttreatment of results [18]. In this process, by building the models, the simulation results could indicate the distribution graph of visibility of fire smoke in the model [19].

*2.1. Establishment of an Abstract Model.* After the investigation and the analysis of the existing super high-rise office buildings in China, this study confined the research subjects to those with a height of at least 150 m, a typical floor plane

of less than 3,000 m<sup>2</sup>, and a tube-in-tube structure. To carry out a comparative simulation experiment, the plane prototype of the typical floor of an office building was made. The plane was a 45 m × 45 m square plane, with an area of 2,475 m<sup>2</sup>. In the center of the plane was a 17 m × 17 m core tube. The space beyond the core tube was an integrated office area of 2,186 m<sup>2</sup>, whose story height was 4.2 meters and net height was 2.9 meters.

In most of the super high-rise office building designs, the atrium is the primary spatial element of plane design. The position and form of atrium is the main factor leading to different typical floor planes. Hence, by incorporating an atrium into the plane prototype, it can be deformed and derived. The approach and logic of incorporating the atrium have two aspects worth discussion: (1) In a fire hazard, the impact of the atrium area on visibility of the office zone; (2) in a fire hazard, the influence of position of atrium space in the office floor plane on visibility of office space. Responding to the above questions, (3) by gradually increasing the area of the atrium in the plane prototype, the derivative plane is obtained, which is shown in Figure 1. (4) By changing the position of the atrium in the plane prototype, the derivative plane is obtained, which is shown in Figure 2. Lastly, six typical floor planes (plane A–plane F) are concluded to serve as the typical floors and planes in this research.

In the above six typical planes, other than the core tube, all areas of plane A are the office zone, which is distributed in the way similar to the character shape “回.” It has the highest efficiency and stable structure for office utilization. However, due to the excessive spatial continuity, the horizontal smoke spreading rate is rather high. There is no atrium space to contain the excessive depth of the office zone, so the smoke exhaust efficiency is low, and crowd evacuation is difficult to be carried out. In plane B/C, plane D, and plane E/F, when the atrium space ratio increased, the office zone area decreased, and the smoke exhaust efficiency increased as well.

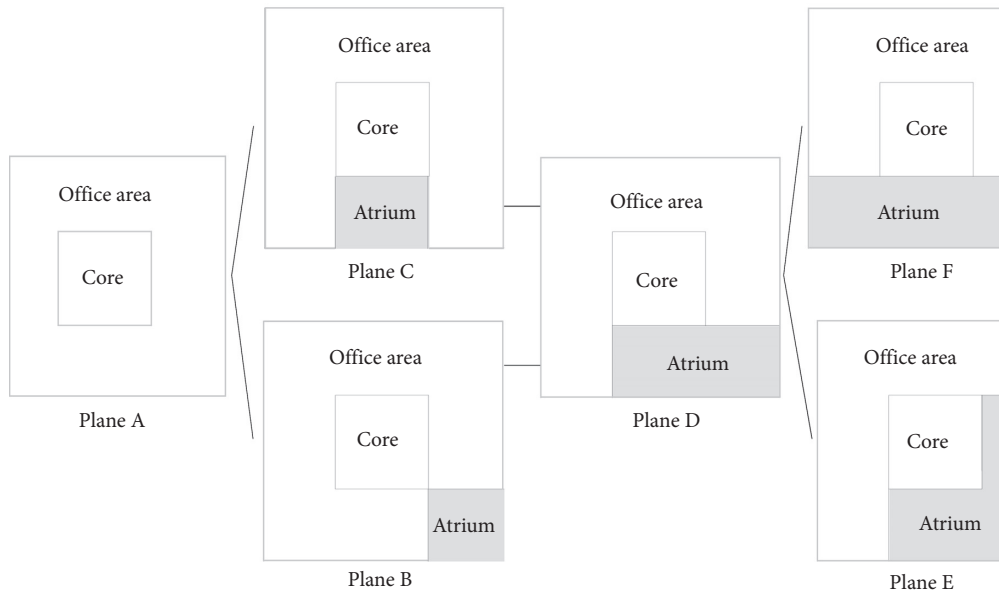


FIGURE 1: Different area proportions of the same atrium position in the plane prototype.

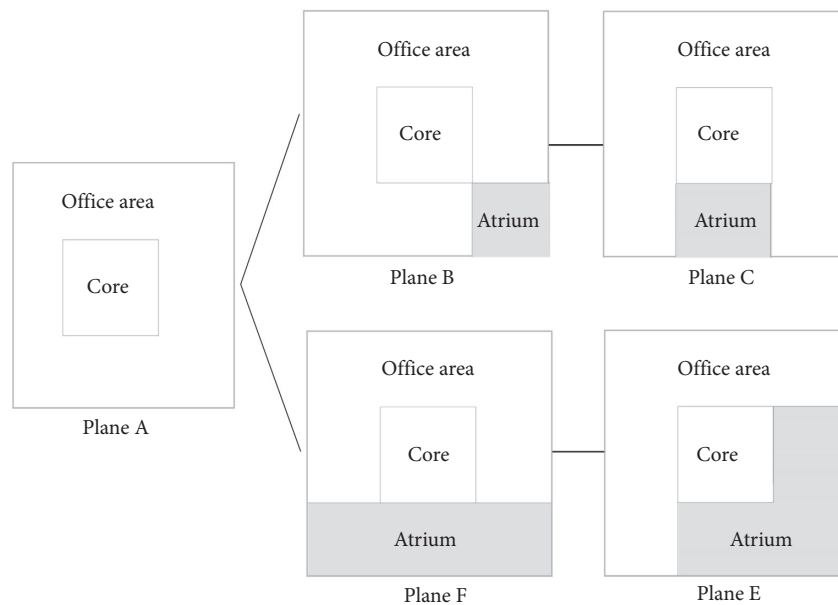


FIGURE 2: Different positions of the same atrium area in the plane prototype.

However, when the area of the office zone became smaller, the time for smoke to sink declines, which may negatively influence visibility. Moreover, smaller office zone reduced the service efficiency and structural stability of buildings. Plane B/C and plane E/F are the typical floor planes with the same atrium area. The atrium positions of the two planes are, respectively, in the single-sided area and corner area of the “回”-shaped office zone. It significantly improved the plane modality of the office area and affected the routes of smoke

spreading. The contrast and analysis of plane features of the six typical floors are shown in Table 4.

In the comparative simulation experience, a piece of virtual land was chosen for a single super high-rise building with a  $45\text{ m} \times 45\text{ m}$  base. The six typical planes were modeled according to the same rules, and their planes were initially designed according to the existing standards to ensure that all the six buildings were a super high-rise office building 294 m high. Besides, there was no atrium in Model A, while

TABLE 4: Contrast of spatial elements in the six typical floor planes.

| Typical planes                         | Plane A | Plane B     | Plane C           | Plane D  | Plane E     | Plane F           |
|--|---------|-------------|-------------------|----------|-------------|-------------------|
| Atrium area ratio                      | 0       | 1/9         | 1/9               | 2/9      | 1/3         | 1/3               |
| Atrium position                        | —       | Corner area | Single-sided area | —        | Corner area | Single-sided area |
| Inner side of the atrium               | —       | 2 planes    | 3 planes          | 3 planes | 4 planes    | 3 planes          |
| Utilization efficiency of office space |         | High        |                   | Medium   |             | Low               |

there was a 42 meter-high atrium for every ten floors in Models B to F. The comparative models are shown in Figure 3, where light pink indicates the scope of simulation and dark red represents the floor with fire. Geometric models of simulation scope are shown in Figure 4.

**2.2. Design of Fire Scenes.** According to the standards, geometric models were established based on the hypothesized comparative simulation schemes, and the fire scenes were designed. In all the six fire scenes, the  $t^2$  fire was regarded as the source of fire, and the heat release rate was medium.

To ensure that people on the typical floor of super high-rise office buildings could evacuate in two different directions, the planes were optimized in the geometric model designs: a 2-meter-wide corridor was created adjacent to the core tube for the two office areas separated by the atrium in the planes equipped with an atrium, so that the office areas always encircle the core tube. As the whole typical floor was an integrated office space, the office area was not furnished with a fire shield. However, the fire shield was utilized to separate the atrium from the office area, and the atrium was established as a separate fire safety zone. An adequate number of smoke safety zones were established in all the six scenes. The largest smoke safety zone was 434 m<sup>2</sup>, smaller than the maximum of 500 m<sup>2</sup> as specified in the standards. The height of all the antismoke boards was 0.6 m, higher than the maximum of 0.5 m as specified in the standards.

As for the layout of the firefighting facilities, the sensing temperature of all the automatic sprinklers was 68°C, and the wind speed of the smoke discharge outlets was 5 m/s, and the outlets would stop working if the temperature reached 280°C. At least two smoke discharge outlets were installed in each smoke safety zone in the office area, and the distance between outlets was shorter than 60 m. The specification of smoke discharge outlet was determined according to the area of the smoke safety zone. According to previously mentioned standard requirements and formulas, ten 0.6 m × 0.9 m outlets were installed. The lower parts around the office area were equipped with five air supplies, and the wind speed was 5 m/s. Besides, the air supply amount met the minimum value (no less than 50% of smoke discharge amount) as specified in the national standards. In the atrium area, the atrium worked as a separate fire safety zone. As the height of the atrium was over 12 meters, the mechanical smoke discharge was adopted. All the smoke discharge outlets were installed on the top of the atrium and on the sides of each structural floor. Smoke shields were installed between the atrium and the office area and would be shut 30 s after the smoke alarm went off. A hole sharing the same

size with mesh was used to simulate a crack. A controlling device was installed on the curtain wall closest to the burner. If the temperature reached 500°C (flash burning), the window would be broken (vanish). Horizontal slices were installed 1.5 m above the ground to collect data of visibility [20].

The floor plans of the geometric models and the firefighting facilities are shown in Figure 5. The parameters of each fire scene are shown in Table 5.

### 3. Results and Discussion

Multiple slices were installed for measurement. Holes sharing the same size with a mesh cell were distributed evenly around the model to simulate the cracks between a normal curtain wall and the floor slab and simulate a real scene. A comparative experiment on the six models was carried out. In a fire, the basic logic of smoke control was as follows. In Models A to F, the burner started to burn from second 0, and the first sprinkler was initiated at a different time, and the HRR curve was kept constant; at this moment, the fire continued to burn but stopped spreading (for specific data, see Figure 6). After the first smoke alarm was initiated, the exhaust system was initiated, and the smoke shield was shut 30 seconds later. In Models B to F, thermocouples were placed near the window closest to the burner and beside the glass in the atrium garden. The controlling device was installed, and when the temperature reached 500°C (flash burning), the window would be broken (vanished). None of the Models B to F reached 500°C.

**3.1. Results of Visibility Simulation.** Sections at different temporal points of the horizontal plane 1.5 m (average practical height) above the ground were taken to observe the changes of the visibility graph, in order to obtain the results of simulation in six fire scenes (Table 6).

The results of visibility simulation show the change to visibility of all the typical planes under the same fire condition. ① In Fire Scene 1, there was no atrium, and the office area encircled the core tube. In a fire, the visibility in the fire safety zone of the fire source was reduced to below 10 m 120 s after the fire started. Then, smoke spread to both sides at the same time, and the visibility of two adjacent fire safety zones began to decline and then went down to less than 10 m 174 s after the fire started. When the smoke spread for 230 s, the visibility of the whole office area was less than 10 m. ② In Fire Scene 2, there was an atrium on the typical floor. As the atrium was located at one corner of the plane, the core tube was still encircled by the office area. In a fire, the visibility in the fire safety zone of the fire source declined to less than

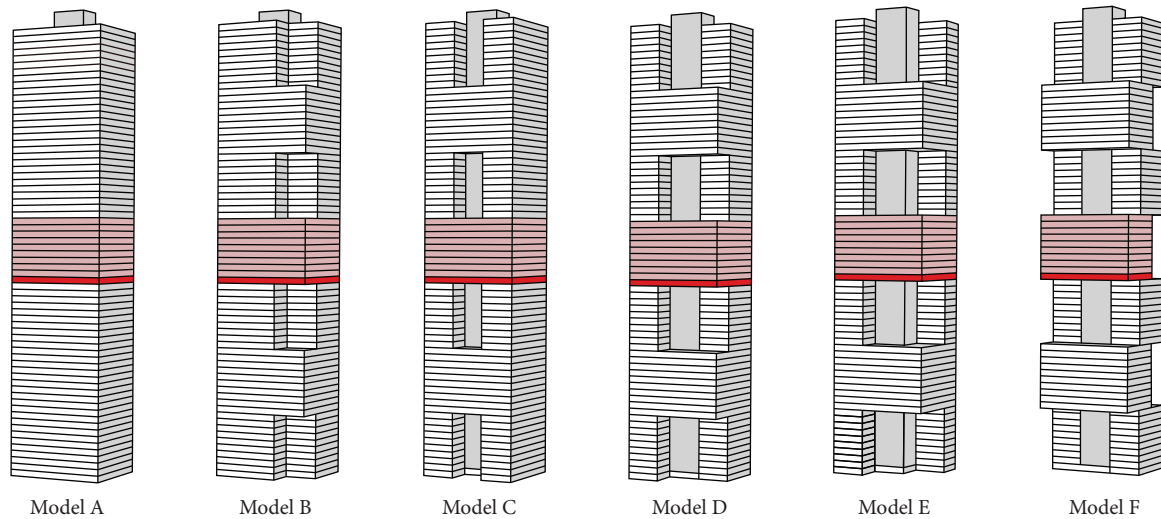


FIGURE 3: Architectural models of the typical planes.

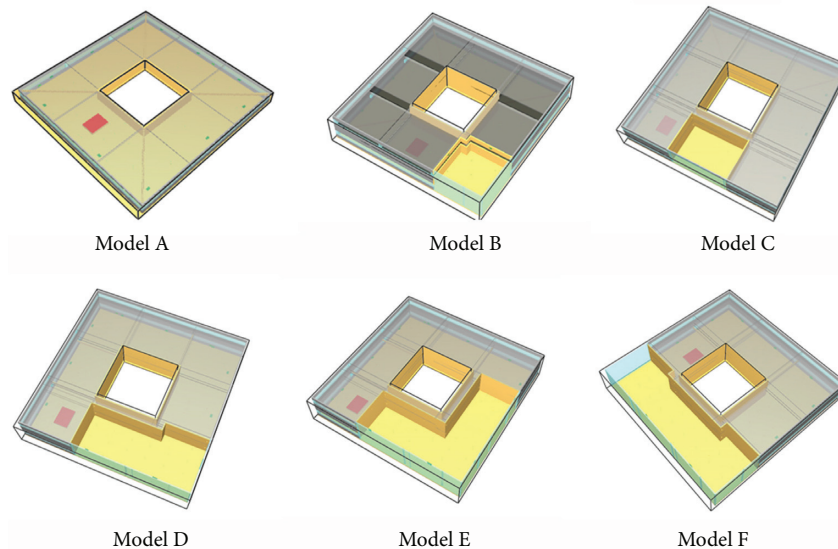


FIGURE 4: Geometric models.

10 m 96 s after the fire started. Then, smoke spread to both sides at the same time. As the size of two openings between the fire area and the two adjacent areas was different, there was a noticeable asymmetric distribution of the reduction rates of visibility: the visibility of the area with the larger opening was reduced significantly, while that with the smaller opening dropped slower. When the smoke spread for 255 s, the visibility of the whole office area was less than 10 m. ③ In Fire Scene 3, there was an atrium on the typical floor. The atrium covered one-fourth of the circumference of the core tube, and the office area covered three-fourths of the circumference. In a fire, the visibility in the fire safety zone of the fire source was reduced to less than 10 m 78 s after the fire started. Then, smoke spread to both sides at the same time. As the size of two openings between the fire area and the two adjacent areas was different, there was a noticeable

asymmetric distribution of the reduction rates of visibility: the visibility of the area with the larger opening was reduced significantly, while that with the smaller opening dropped slower; besides, a visual corridor with high visibility was formed between the atrium and the core tube. The visibility of the whole office zone was below 10 m 1000 s after the fire started. ④ Fire Scene 4 had a similar process of visibility reduction with Fire Scene 3. The atrium still covered one-fourth of the circumference of the core tube. But, as the atrium area was larger and the office area was smaller in Fire Scene 4 than Fire Scene 3, it took a shorter time (260 s) for the visibility of Fire Scene 4 to drop to less than 10 m. ⑤ In Fire Scene 5, the atrium on the typical floor continued to become larger, and the office area continued to shrink to 1,136 m<sup>2</sup>, and both the atrium and the office area covered half of the circumference of the core tube, respectively. In a

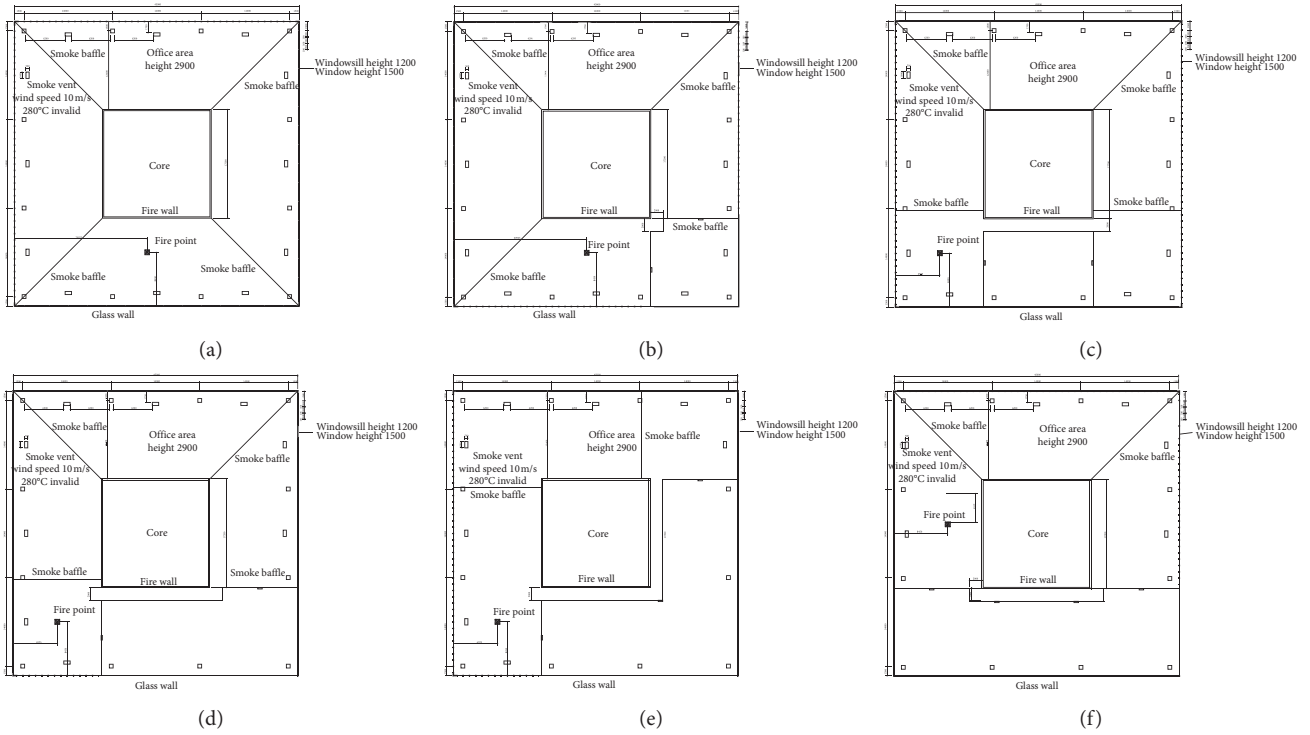


FIGURE 5: Floor plans of the comparative models.

TABLE 5: Parameters of the fire scene design.

|   | A  | B    | C    | D    | E    | F    |
|---|--|------|------|------|------|------|
| Office area ( $m^2$ )   | 1736   | 1536 | 1536 | 1336 | 1136 | 1136 |
| Atrium area ( $m^2$ )   | 0  | 200  | 200  | 400  | 600  | 600  |
| The amount of smoke discharged from the office area ( $m^3/s$ ) | 28.9   | 25.6 | 25.6 | 22.3 | 22.3 | 22.3 |
| The amount of smoke discharged from the atrium area ( $m^3/s$ ) | 0  | 4    | 4    | 7    | 10   | 10   |
| Height of smoke shield (mm)                                     | 600  |      |      |      |      |      |
| Fuel type   | Built-in POLYURETHANE-PyroSim $C = 6.3$ , $H = 7.1$ , $O = 2.1$ , and $N = 1.0$ ; common data for simulating office fire           |      |      |      |      |      |
| Soot yield, $y$   | 0.07 kg/kg fuel  |      |      |      |      |      |
| CO yield, $y_{CO}$  | 0.04 kg/kg fuel  |      |      |      |      |      |
| Heat of combustion, $\Delta H_c$                                | $1.3 \times 10^4$ kJ/kg, $O_2$   |      |      |      |      |      |
| Radioactive fraction  | 35%  |      |      |      |      |      |
| Burner height (mm)  | 500  |      |      |      |      |      |
| Fire growth rate  | $Q = 0.017 t^2$ medium fire  |      |      |      |      |      |
| HRRPUA  | 1000 kw/m <sup>2</sup> (FDS), burner size of 5 m × 4 m   |      |      |      |      |      |
| Peak heat release rate  | 20 MW or HRR controlled by spray   |      |      |      |      |      |
| Fire sprinkler  | Standard sprinkler head: RTI 135, temperature 68°C, C factor -0.85, spacing of 3.6 m, referring to the Chinese standard for office |      |      |      |      |      |
| Smoke detection parameters                                      | Photoelectric type, spacing 5.8 m, referring to Chinese standard.  |      |      |      |      |      |

fire, the visibility in the fire safety zone of the fire source was reduced to less than 10 m 85 s after the fire started. As the size of two openings between the fire area and the two adjacent areas was different, there was a noticeable asymmetric distribution of the reduction rates of visibility: the visibility of the area with the larger opening was reduced significantly, while that with the smaller opening dropped slower; besides, a visual corridor with high visibility was formed between the atrium and the core tube. When the smoke spread for 230 s, the visibility of the whole office area was less than 10 m. © Fire Scene 6 shared the same area of

the office zone (1,136 m<sup>2</sup>) with Fire Scene 5. The atrium took the shape of “—,” covering one-fourth of the circumference of the core tube. In a fire, the visibility in the fire safety zone of the fire source was reduced to less than 10 m 85 s after the fire started. As the size of two openings between the fire area and the two adjacent areas was different, there was a noticeable asymmetric distribution of the reduction rates of visibility: the visibility of the area with the larger opening was reduced significantly, while that with the smaller opening dropped slower; besides, a visual corridor with high visibility was formed between the atrium and the core tube. When the

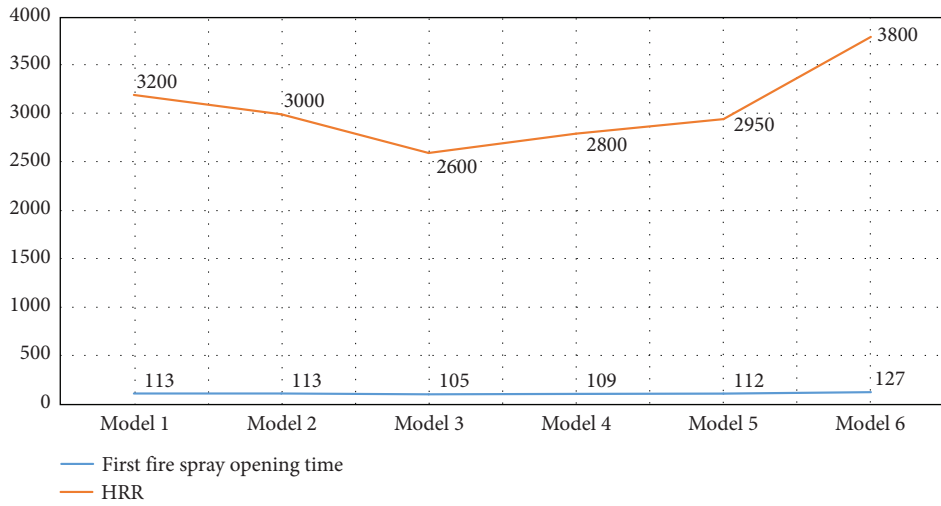


FIGURE 6: Comparison of data about smoke control.

TABLE 6: Visibility distribution graph of different temporal points in six fire scenes.

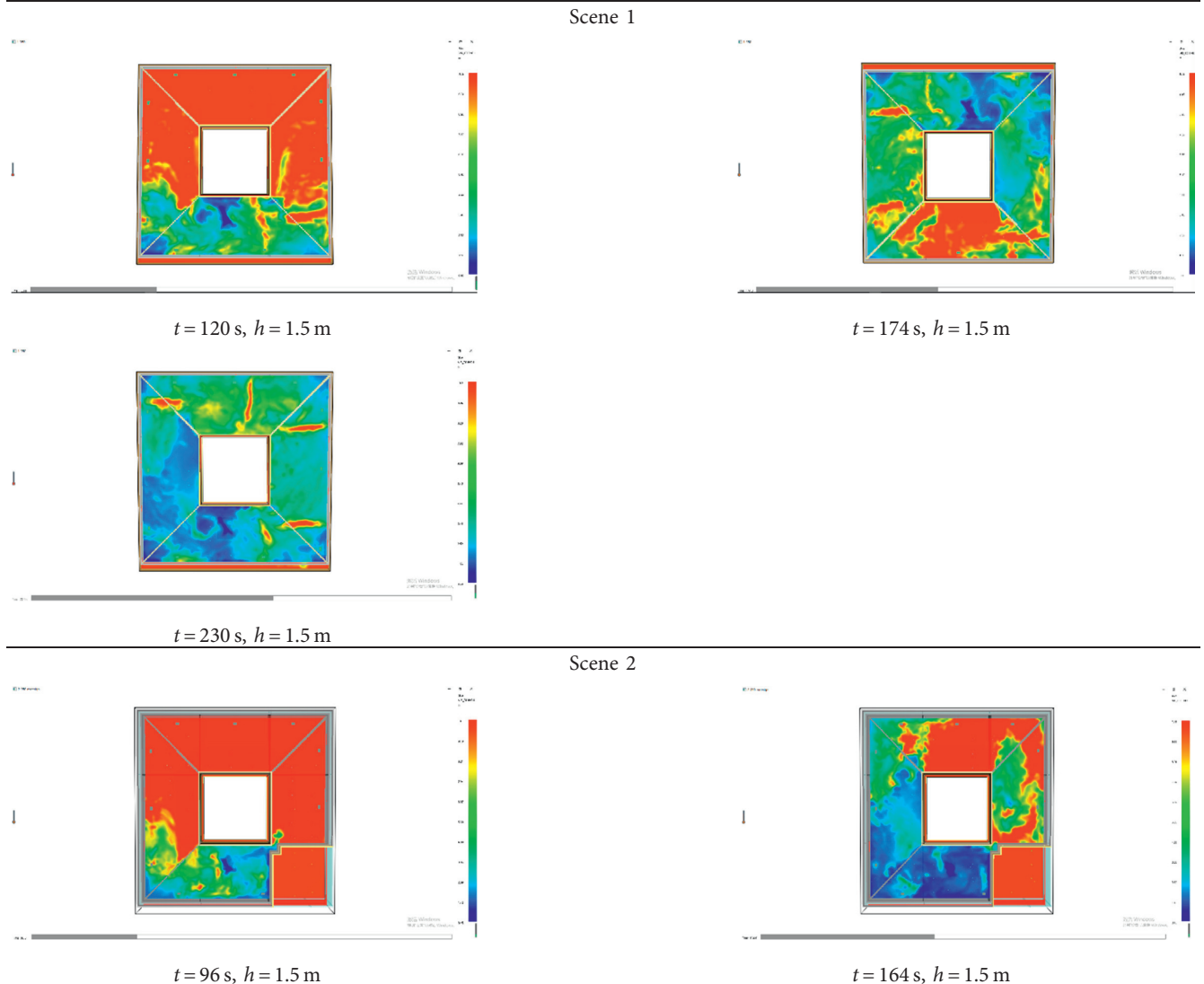


TABLE 6: Continued.

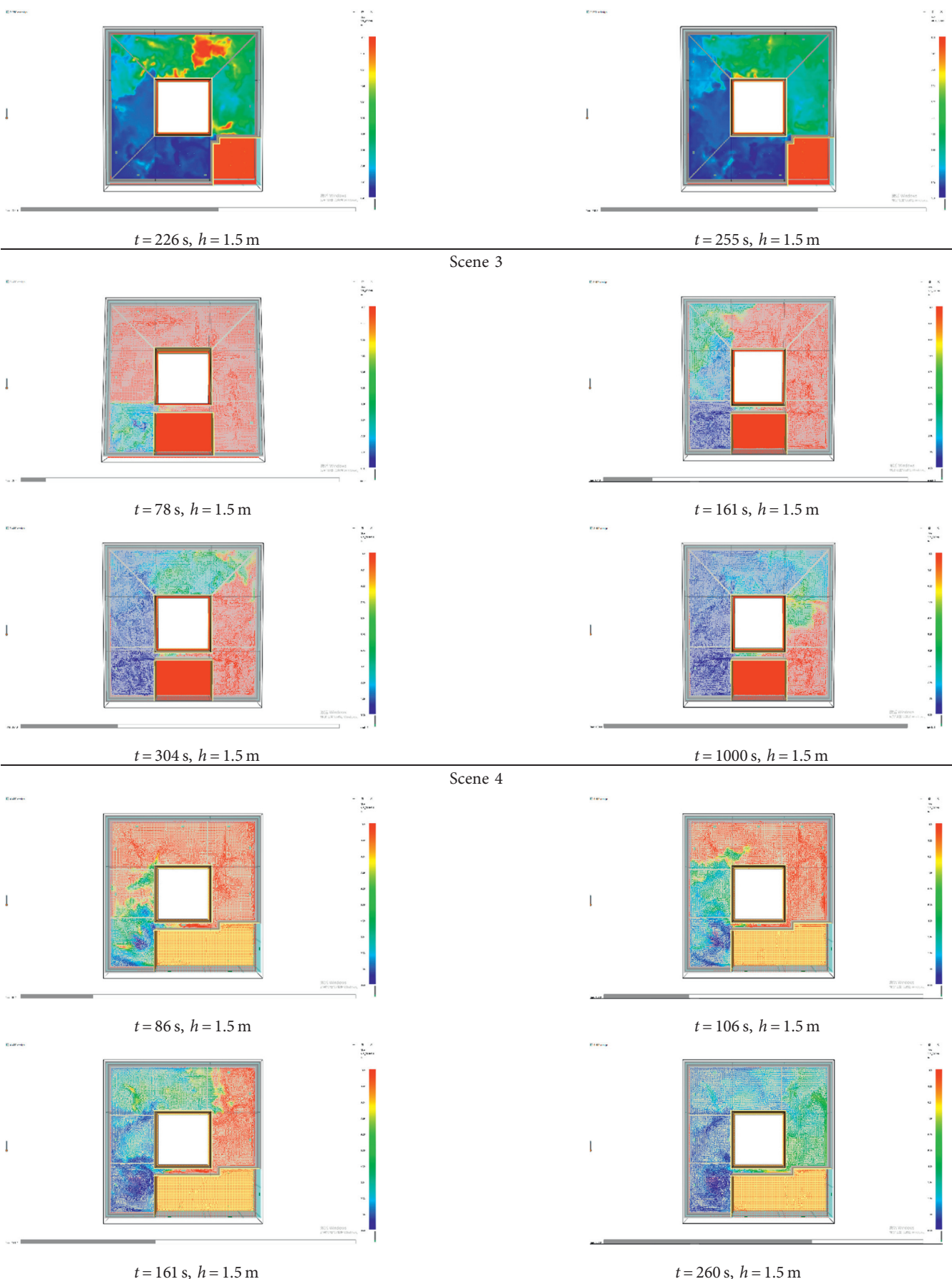
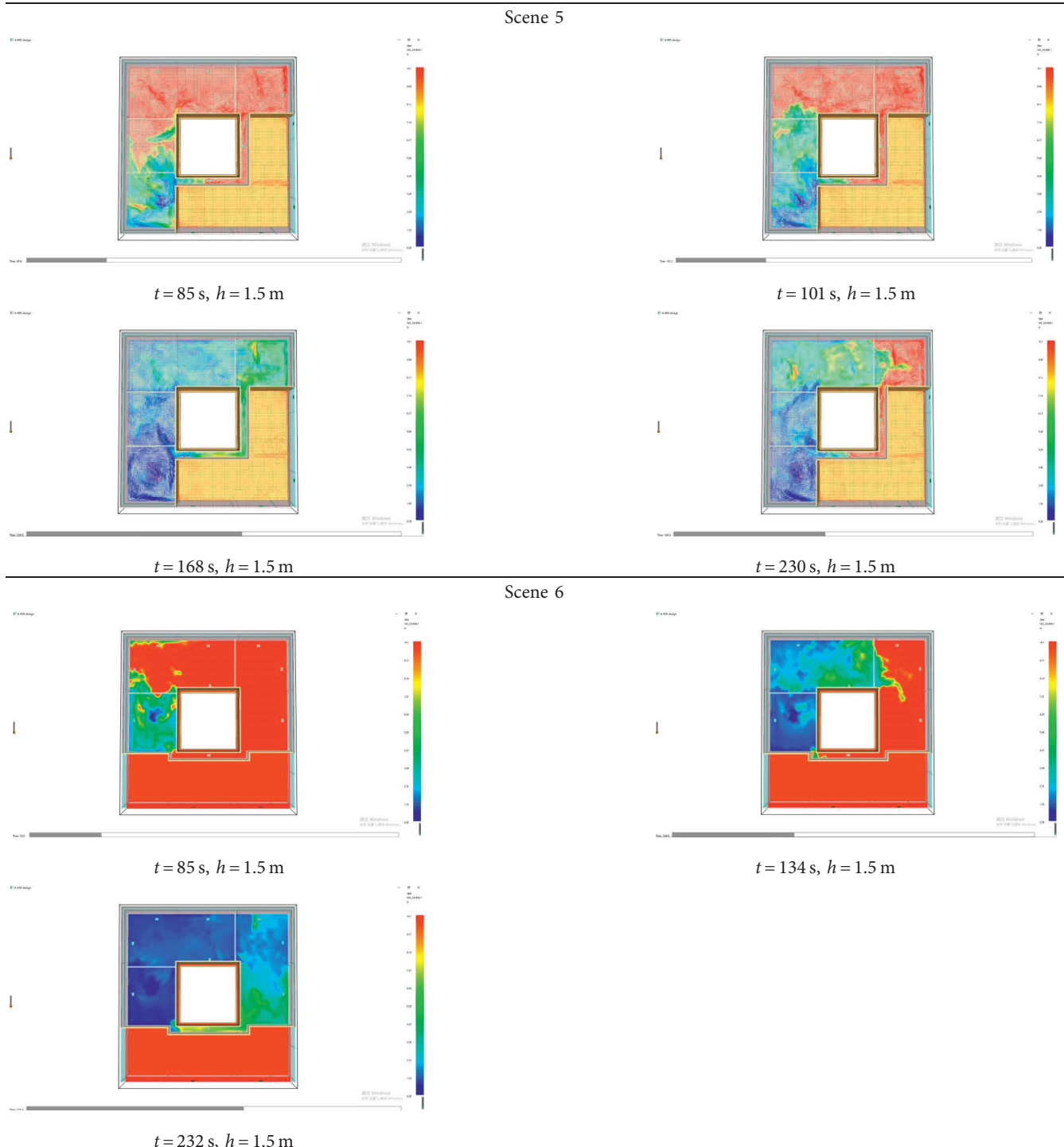


TABLE 6: Continued.



smoke spread for 232 s, the visibility of the whole office area was less than 10 m.

### 3.2. Visibility Comparison Analysis

3.2.1. Overall Visibility Comparison Analysis. When the area of the typical floor and the height of the floor remained

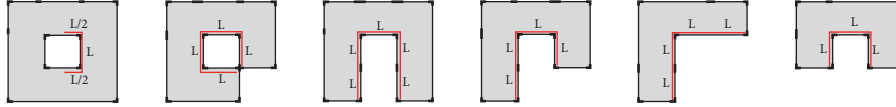
unchanged, the area (volume) of the office zone gradually became smaller with the increasing single-floor area of the atrium from Fire Scene 1 to Fire Scene 6, but the time for the visibility to drop to 1.5 m due to the spreading of smoke did not show any sign of becoming shorter. The duration of visibility reduction reached a peak value in Fire Scene 3. In other words, when the office area was 1,536 m<sup>2</sup> and the atrium was established on one side of the core tube, the



TABLE 7: Comparison of visibility data among different fire scenes.

|  | Scene 1 | Scene 2 | Scene 3 | Scene 4 | Scene 5 | Scene 6 |
|--|---------|---------|---------|---------|---------|---------|
| Office area (m <sup>2</sup> )            | 1736    | 1536    | 1536    | 1336    | 1136    | 1136    |
| Smoke spreading path                     | 2L      | 4L      | 5L      | 4L      | 4L      | 3L      |
| Duration of overall visibility reduction | 230 s   | 226 s   | >1000 s | 260 s   | 230 s   | 232 s   |

Plane shape



duration of visibility reduction was the longest. Under other circumstances, the area (volume) of the office zone, the area of the atrium, and the location of the atrium all had influence on the duration of overall visibility reduction of the office area. Therefore, a comparative research on the six models under two circumstances “when the area (volume) of the office zone was different” and “when the area (volume) of the office zone was the same but the location of the atrium was different” was carried out.

- (1) When the area (volume) of the office zone was different but the location of the atrium was the same.

In the comparison among Fire Scenes 1, 2, and 4, the office area of Models A, B, and D decreased progressively with a tolerance of 200 m<sup>2</sup>, and the spreading rate of smoke showed a trend of increasing progressively ( $V_A > V_B > V_D$ ). But, as there was a great difference in the plane of the office area of the three models, there was some difference in the smoke spreading path. According to the abstract plane of the office area (Table 7), the smoke in Model A spread to both sides in a symmetric way, so its smoke spreading path was counted as “2L.” The smoke spreading path of Model B and Model D was a one-way path which was counted as “4L.” The relationship among the three smoke spreading paths of the three models was “ $LA < LB = LC$ .” According to the simulation results, the total duration of visibility reduction of the three fire scenes was almost the same. This demonstrates that the establishment of the atrium reduced the area of the office zone, but this improved the smoke spreading path, and a one-way smoke spreading path had greater impact on overall visibility than the area (volume) of the office zone.

- (2) When the area (volume) of the office zone was the same but the location of the atrium was different.




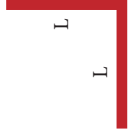

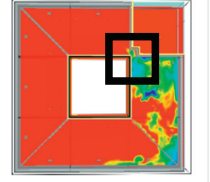
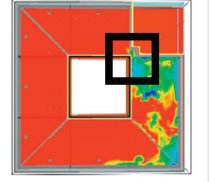
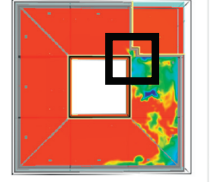
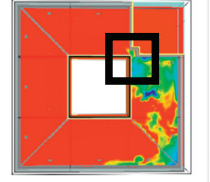
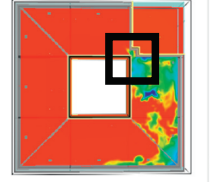
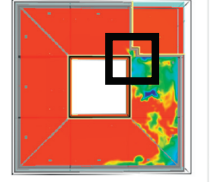
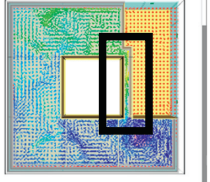
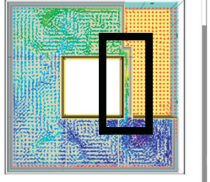
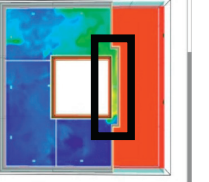

In the comparison between Fire Scene 2 and Fire Scene 3, the office area (volume) and smoke spreading rate of Models B and C were the same. According to the abstract plane of the office area (Table 7), the smoke spreading path of Model B was “4L” and that of Model C was “5L.” The simulation

results show that the duration of visibility reduction in Fire Scene 3 was remarkably longer than that in Fire Scene 2.

In the comparison between Fire Scene 5 and Fire Scene 6, the office area (volume) and smoke spreading rate of Models E and F were the same. According to the abstract plane of the office area (Table 7), the smoke spreading path of Model E was “4L” and that of Model F was “3L.” The simulation results show that the duration of visibility reduction in Fire Scene 5 was longer than that in Fire Scene 6. It is obvious that the smoke spreading path had greater influence on the overall visibility of the office than smoke spreading rate. In other words, the location of the atrium had greater influence on the duration of overall visibility reduction than the area (volume) of the office zone.

**3.2.2. Corridor Visibility Comparison.** According to the standards, an evacuation exit was set on one side of the office area adjacent to the atrium to ensure that each fire safety zone was equipped with two evacuation exits, and the exit must be as close to the core tube as possible. Therefore, evacuation corridors of different lengths between the atrium and the core tube were established according to different floor planes. As there was no atrium in Model A, there was no evacuation corridor. In Model B, the atrium was established at one corner of the plane, so an L-shaped corridor (represented as “L”) was formed between the atrium and the core tube. In Model C, a corridor with the length of “L” was formed between the atrium and the core tube. The two ends of the corridor were connected with two fire safety zones. In Model D, a corridor with the length of “L” was formed between the atrium and the core tube. One end of the corridor was connected with the fire safety zone, and the other end was linked with the L-shaped corner of the fire safety zone. In Model E, a corridor with the length of “2L” was formed between the atrium and the core tube. The two ends of the corridor were connected with two fire safety zones. In Model F, a corridor with the length of “L” was formed between the atrium and the core tube. One end of the corridor was connected with the fire safety zone, and the other end was linked with the L-shaped corner of the fire

TABLE 8: Comparison of overall visibility data among the fire scenes.

|                                      | Scene 1   | Scene 2  | Scene 3   | Scene 4   | Scene 5   | Scene 6   |
|--------------------------------------|---|--|---|---|---|---|
| Office area                          | 1736 m <sup>2</sup>   | 1536 m <sup>2</sup>  | 1536 m <sup>2</sup>   | 1336 m <sup>2</sup>   | 1136 m <sup>2</sup>   | 1136 m <sup>2</sup>   |
| Passage length                       | —   | 0  | L   | L   | 2L  | L   |
| Port type                            | —   | I  | 0   | I   | 0   | 2 I   |
| Passage plane                        | —   |   |   |    |      |    |
| Duration of visibility reduction (s) | —   | 53 s~77 s  | > 53 s~1000 s   | 54 s~260 s  | 54 s~230 s  | 53 s~232 s  |
|                                      |  |   |  |  |    |  |
|                                      |   | Scene 2  |   |   | Scene 3   |   |
|                                      |   |  |   |   |   |   |
|                                      |   |  |   |   | Scene 5   |   |
|                                      |   |  |   |   |  |   |
|                                      |   |  |   |   |   | Scene 6   |
|                                      |   |  |   |   |  |   |

shows the corridor zone between atrium and core tube. It is also the key area leading to differences between personnel evacuation and smoke spreading.

safety zone. Table 8 shows the length and end of different evacuation corridors as well as the duration of corridor visibility reduction.

In the comparison between Fire Scene 2 and Fire Scene 4, there was no evacuation corridor in Fire Scene 2, and the two fire safety zones were connected through the corner only. In Fire Scene 4, there was an evacuation corridor with a length of “L,” and a corner was set on one end. According to the results, the visibility on one end of the corridor began to decline nearly at the same time in both fire scenes, but the duration of visibility reduction in Fire Scene 4 was noticeably longer than that in Fire Scene 2. In the comparison among Fire Scenes 3, 4, and 6, no corner was set on either end of the corridor in fire scene with the corridor length in all the three fire scenes being “L.” But, there were smoke discharge outlets in Fire Scene 3, a corner was set on one end of the corridor in Fire Scene 4, and there were corners on both ends of the corridor in Fire Scene 5. According to the results, there was marked difference in the duration of visibility reduction among the three fire scenes, and the area with a corner effectively lengthened the duration of visibility reduction on the ends of the corridor and in the corridor. Compared with a corner, a smoke discharge outlet in the corridor was more effective in slowing down the visibility reduction rate and saving more effective evacuation time for the people in the corridor.

#### 4. Conclusion

- (1) The location of the atrium on the typical floor has a direct effect on the duration of visibility reduction.

The area (volume) of the office zone is an influencing factor to the smoke spreading rate in a fire. But, when the area of the office zone is the same or almost the same, the duration of visibility reduction in the area is, to a larger extent, influenced by the smoke spreading path. In the floor plan of a typical floor, changing the location or the plane of the atrium would effectively change the smoke spreading path in the office area. If the plane area of the typical floor is the same, more diverse and longer smoke spreading paths will lengthen the overall duration of visibility reduction on the typical floor and thus make it easier for people to evacuate.

Given the structure of super high-rise office buildings, a typical floor without atriums tends to take a symmetric form. In a fire, the smoke spreading path is the same as the evacuation path, and the fire would spread towards both ends of the fire safety zone where the fire started. This will impede evacuation. An atrium can change the smoke spreading path to ensure that the evacuation path is different from the smoke spreading path, which will facilitate the evacuation.

A longer edge shared by the atrium and the office area will lengthen the smoke spreading path and expand the duration of visibility reduction, which will lengthen the evacuation time.

In summary, establishing an atrium on the typical floor of super high-rise office buildings and lengthening the edge shared by the atrium and the office area will effectively lengthen the evacuation time in a fire.

- (2) Establishing evacuation corridors around core tubes can hinder the duration of visibility reduction.

Establishing an evacuation corridor can effectively stop smoke spreading, change the smoke spreading path in the whole office area, and effectively separate the evacuation path from the smoke spreading path. A longer evacuation corridor will lengthen the duration of visibility reduction in the evacuation corridor and will thus facilitate evacuation. Setting corners on both ends of the evacuation corridor can effectively slow down the visibility reduction on both ends and inside the evacuation corridor. But, a corner on the evacuation path would impede the evacuation and is against the design principles of an evacuation corridor. Therefore, corners are not recommended. It is suggested that smoke discharge outlets should be installed in the evacuation corridor.

#### Data Availability

All the data used to support the findings of this study are included within the article.

#### Conflicts of Interest

The author declares that there are no conflicts of interest.

#### Acknowledgments

This study was supported by Optimization Design for Vertical Penetrating Space of High-Rise Based on Performance Fire Protection: A Case of the Guangdong-Hong Kong-Macau Great Bay Area (51908357), National Natural Science Foundation of China (2020.01-2022.12), and Research on Vertical through Space Optimization Design of Super High-Rise Buildings based on Pyrosim Software Simulation (000002110332), Shenzhen University (2019.09-2022.09).

#### References

- [1] A. Cowlard, A. Bittern, and C. Abecassis-Empis, “Fire safety design for tall buildings,” *Procedia Engineering*, vol. 62, pp. 169–181, 2013.
- [2] C. H. Torero, W. L. Liu, X. Y. Zhang, and Z. N. Xiao, “Fire technology development and research focus of super high-rise building,” *Building Science*, vol. 9, pp. 82–88, 2018.
- [3] D. Madrzykowski and W. D. Walton, “Cook county administration building fire, 69 West Washington, Chicago, Illinois: heat release rate experiments and FDS simulations (2002),” 2003.
- [4] K. Knight, *Report to the Secretary of State by the Chief Fire and Rescue Adviser on the Emerging Issues Arising from the Fatal Fire at Lakanal House, Camberwell on 3 July 2009*, Ministry of

- Housing, Communities and Local Government, Camberwell, UK, 2009.
- [5] C. TV Network, *Apologizes for Fire*, New York Times, New York, NY, USA, 2009, <http://www.nytimes.com/2009/02/11/world/asia/11%20beijing.html>.
  - [6] W. Zheng, "Application of computer simulation technology [CST] in buildings performance-based fire protection design," *Procedia Engineering*, vol. 37, pp. 25–30, 2012.
  - [7] Z. Xing, X. Zhao, H. Song, and W. Gao, "Applied research of performance-based fire protection design in a large building," *Procedia Engineering*, vol. 11, pp. 566–574, 2011.
  - [8] W. Yao, H. Huang, S. Shen, L. Qiao, and W. Zhang, "Fire risk mapping based assessment method applied in performance based design," *Fire Safety Journal*, vol. 56, pp. 81–89, 2013.
  - [9] Y. Li, *Performance-Based Building Design*, Chemical Industry Press, Beijing, China, 2005.
  - [10] R. L. P. Custer, *SFPE Engineering Guide to Performance Based Fire Protection Analysis and Design of Buildings*, Natl Fire Protection Assn, MA, USA, 2000.
  - [11] F. F. Dong, "Application of performance-based fire prevention evaluation technique in large space building," *Architecture Technology*, vol. 2, pp. 215–217, 2017.
  - [12] Y. X. Lei, C. K. Deng, J. Ma et al., "Numerical simulation of fire smoke spread in a super high-rise building for different fire scenarios," *Advances in Civil Engineering*, vol. 2019, pp. 1–11, Article ID 1659325, 2019.
  - [13] Z. C. Wang and X. Liu, "Analysis on the evacuation from a super high-rise building," *Fire Science and Technology*, vol. 11, pp. 1524–1526, 2018.
  - [14] M. Rodrigo and S. Marshall, "The development of a real performance-based solution through the use of People Movement Modelling Analysis (PeMMA) combined with fire modelling analysis," *Safety Science*, vol. 50, pp. 1485–1489, 2012.
  - [15] B. Liu, "The application of "sub-safe area" in the performance-based fire protection design," *Procedia Engineering*, vol. 11, pp. 280–287, 2011.
  - [16] T. Zhang, "The research in optimization fire protection design of high-rise complex space in view of performance-based design methods," *Development of Performance-based Fire Code*, vol. 35, no. 10, pp. 54–55, 2016.
  - [17] T. Zhang and J. Zeng, "Design optimization of fire performance of atrium space," *Telkommnika*, vol. 14, pp. 125–132, 2016.
  - [18] C. Xing and S. Lv, "Research on complex mine fire smoke intelligent control based on Pyrosim." *Code for Fire Protection Design of Buildings (GB 50016-2014)*, vol. 10, pp. 9–10, 2012.
  - [19] A. B. O Nja, "Concept of validation in performance-based fire safety engineering," *Safety Science*, vol. 52, pp. 57–64, 2013.
  - [20] L. Geyer Bellamy, R. Max-Lino, Z. Harrison Bahrami, and B. Modha, "An evaluation of the effectiveness of the components of informative fire warning system," in *Safety in the Built Environment, E and FnSpon*, J. Sime, Ed., Springer, Berlin, Germany, 1988.

## Research Article

# Simulation Study on the Spatiotemporal Difference of Complex Neurodynamics between P3a and P3b

**Xin Wei,<sup>1</sup> Xiaoli Ni ,<sup>1</sup> Junye Liu,<sup>2</sup> Haiyang Lang,<sup>2</sup> Rui Zhao,<sup>3</sup> Tian Dai,<sup>3</sup> Wei Qin,<sup>3</sup> Wei Jia,<sup>4</sup> and Peng Fang<sup>5</sup>**

<sup>1</sup>*Institute of Social Psychology, School of Humanities and Social Sciences, Xi'an Jiaotong University, Xi'an 710049, China*

<sup>2</sup>*Teaching and Research Section of Radiation Medicine, Department of Military Preventive Medicine, The Fourth Military Medical University, Xi'an, China*

<sup>3</sup>*Sleep and Neuroimage Group, School of Life Sciences and Technology, Xidian University, Xi'an 710126, China*

<sup>4</sup>*Xi'an Gaoxin No. 2 Middle School, Xi'an 710119, China*

<sup>5</sup>*Department of Military Medical Psychology, Air Force Medical University, Xian 710032, China*

Correspondence should be addressed to Xiaoli Ni; [nixiaoli@xjtu.edu.cn](mailto:nixiaoli@xjtu.edu.cn)

Received 27 March 2020; Revised 1 June 2020; Accepted 22 June 2020; Published 24 December 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Xin Wei et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The integration of event-related potential (ERP) and functional magnetic resonance imaging (fMRI) helps to obtain and study neural networks with high temporal and spatial resolution. EEG/fMRI data proves that in the visual tritestimulus oddball paradigm, two P300 potentials (P3a and P3b) induced by target stimulation and novel stimulation are detected at the frontal-middle (Fz), center (Cz), and mid-apical (Pz) electrodes. Previous studies have shown that P3a and P3b have different spatial distributions of brain activation, but it is unclear whether they have the same neural mechanism. The purpose of this study is to determine the neuropsychological mechanisms of P3a and P3b, as well as the spatiotemporal differences in neurodynamics between the two ERP subcomponents. In a group of 25 subjects, P300 ERP induced by target stimulation and novel stimulation can be detected at the Fz, Cz, and Pz electrodes. At Cz and Fz, compared with P3b related to the target stimulus, the P3a related to the novel stimulus has a higher amplitude and the waveform declines more slowly. But at Pz, P3b has a higher amplitude than P3a. P3a appeared earlier than P3b at Cz and Fz, but the opposite phenomenon was observed at the Pz electrode. The activated brain regions of P3a included the left frontal-parietal lobe region, left anterior wedge lobe region, and right insula, while the target-driven P3b was significantly associated with BOLD changes in the bilateral fusiform gyrus, the left frontal region, and the bilateral insula. The results showed that the integration of the spatial and temporal information of the two imaging modes, namely, ERP and fMRI, proves the existence of the different brain function processes of the two P300 subcomponents. Through the analysis of the composition of P300, the results further proved that the top-down and bottom-up processing processes have played a role in the occurrence of attention capture. It is just that the modulation effects of the two processing mechanisms are different in different tasks. Therefore, it should be noted that the captured neural mechanism is not a single top-down or bottom-up processing process but should be the result of the interaction between the two.

## 1. Introduction

For the human, rapidly distinguishing the relevant targets from distracting stimuli is one of the critical ability to execute an appropriate response according to the environmental requirement. One widely used task to assess this cognitive ability is oddball paradigms. In a three-stimulus oddball, it presents infrequent target stimuli in a background

of frequent standard stimuli and infrequent novel stimuli, and the subject is instructed to respond mentally or physically to the target stimuli as quickly and accurately as possible, but not to others [1].

The brain activity related to this ability has been widely studied with electrical recordings using ERP. Previous ERP research has shown that the P300 is the characteristic endogenous positive wave, arising between 260 and 500 ms

after the onset of each rare stimulus, which is more prominent over the centroparietal region of the scalp [2]. The oddball tasks can typically evoke two highly characteristic P300 ERPs, the target-related P3b and the novel-related P3a. The P3b is seen as a long positive slow wave occurring after a button press, which is affected by the expected versus actual action effects [3]. And the scalp distribution is defined as the amplitude change over the three midline electrodes (Fz, Cz, and Pz), which typically increases from the frontal to parietal sites and is the largest over parietal areas [4]. An infrequent novel stimulation, presented in a series of frequent standard stimuli without a target stimulation, can produce a positive-going waveform with a frontoparietal maximum amplitude distribution and relatively short peak latency and was dubbed the P3a to distinguish it from the target-related P3b potential [5, 6]. Many studies have found that P3a and P3b have different topographic amplitude distributions and peak latency. P3a and P3b data, using a difficult target/standard discrimination task in a three-stimulus paradigm from 120 healthy young adults, have shown that P3a has a central maximum, whereas P3b has a parietal maximum, and the peak latency for both was shorter over the frontal and longer over the parietal sites [7]. The topographic variation suggested that there may be overlapping neural activation patterns, which are functionally distinct [8, 9].

However, EEG traceability positioning is a relatively complicated problem, and there are infinitely many solutions that satisfy the EEG signal recorded by the scalp. Therefore, we need a priori assumption to constrain the solution space to obtain a unique solution. Functional magnetic imaging (fMRI) can be used to study the source localization of P300; many studies have identified brain activation evoked by detecting targets and novels in oddball tasks, which have shown different neuronal networks activated by targets (right temporoparietal junction and inferior frontal gyrus) and by novels (intraparietal sulcus and left inferior frontal sulcus) [10–13]. The relationship between abnormal ERP components and dysfunctional fMRI activation during oddball tasks in patients has also been extensively studied. Patients with frontal lobe lesions demonstrated diminution of P3a amplitude, whereas the same patients demonstrated a parietal maximum for the P3b, which suggested that frontal lobe integrity is necessary for P3a generation [14]. For the patients with comorbid depressive and anxiety disorders, findings showed an abnormal frontal-greater-than-parietal P3b topography in the right hemisphere and the highest P3a amplitude at frontal and central sites at the scalp midline [15]. It is proposed that P300 amplitude is affected by temporal-parietal junction integrity for its absence greatly reduces component size over the parietal area [16, 17]. This connection implies that the P3a and P3b indicate a circuit pathway between frontal and temporal/parietal brain areas [18, 19].

Event-related potential (ERP) and functional magnetic resonance (fMRI) are two advanced functional detection techniques for nondamaging brains today. EEG signal is a biological electrical activity spontaneously generated by the central nervous system, contains a wealth of information on

the state and changes of the nervous system, and can reflect the brain's skin reflection on stimulation. EEG signals are easier to obtain, and it is easier to establish interfaces with external devices (such as computers) to achieve direct interaction between the human brain and computers. This interaction is realized through brain-computer interface technology and then realizes a system to control the brain. EEG/ERP has a high temporal resolution and can track the processing of brain information, but the spatial resolution is low. fMRI can repeatedly observe and study the human brain function (language, cognition, exercise, etc.). It can not only observe the structure and shape of the brain but also get real-time information about blood flow functional images and metabolism. However, fMRI has a high spatial resolution, but a low temporal resolution. Combining EEG and fMRI signals and using information fusion theory to find the relationship between them may become a new and very effective method for scientific analysis and processing of brain function data.

Brain function modeling has always been a big problem that puzzled scholars from various countries. Due to the complexity of brain structure and function, it is difficult to model brain function alone. Internationally, the research on brain function modeling in recent years has mainly focused on the detection of EEG and fMRI signals. In order to improve the sensitivity and accuracy of detection, foreign researchers in recent years have vigorously improved measurement equipment and measurement methods; for example, improving the field strength of fMRI is very high. And some new methods in the extraction of useful signals and noise filtering, feature extraction, and data classification improve the signal quality of EEG and fMRI and the effectiveness of feature extraction, but this progress is slow. A new trend is that brain function detection is developing in conjunction with multiple technologies. For example, MRI and PET use image fusion or registration technology to obtain more information on brain functional activities; at the same time, EEG and MRI signals are obtained for information fusion can get more comprehensive features of brain function and improve resolution. Research in this area has been carried out in Germany, Canada, Japan, and other countries and a few domestic units. The representative of the Charlotte Hospital in Germany scans fMRI images while acquiring EEG signals for pathological research and pattern recognition and localization of brain function when the brain receives different stimuli from the body. But their research has just begun, and the correlation between EEG and fMRI signals is lacking.

Obtaining high imaging resolution and correct interpretation, analysis and processing of brain activity features have become the frontier of current research. The use of brain functional imaging technology to model, classify, and recognize the pattern of human brain nerve response can not only provide a more accurate and rich basis for the diagnosis of certain diseases but also more objectively study the human brain in psychology and intelligence. This aspect of the activity lays the foundation and is a crucial step in further understanding the function of the brain and the connections between the peripheral nerves and the brain. Due to the high

cost and difficulty of using a single method to improve resolution and improve the analysis of brain functional characteristics, and the fact that any single method has its own advantages and disadvantages, no modern medical imaging technology is perfect and can be suitable for all experiments or clinical applications. It is precisely because of the incompleteness of a single imaging method that more and more researchers begin to consider the fusion of multiple modalities of medical images, so it is an effective technical approach to use multiple technical means to obtain composite signals and perform information fusion.

In the present simultaneous ERP/fMRI experiment, 25 healthy participants were confronted with a visual three-stimulus oddball (standard, target, and novel stimuli). The present study aimed to explore the differences of the spatiotemporal neural dynamics between the P3a and P3b and the neuropsychological origins underlying these two ERPs.

The main purpose of this study is to accurately locate the brain activation status of the target recognition process. One of the methods is to use the EEG data collected by the scalp to calculate the activation status in the brain. However, due to the low spatial resolution of the EEG data, the EEG accurate calculation of dipole data also requires certain assumptions, so it is very difficult to accurately obtain the position of the brain source. With the development of modern brain imaging technology, especially the application of functional magnetic resonance technology, it has made an accurate determination of brain activity and location becomes possible. The spatial resolution of functional magnetic resonance technology is very high, but the temporal resolution is limited. In order to further study the brain activation of P300 operation, we combined high-resolution EEG with high-spatial-resolution functional magnetic resonance technology in this process.

To show the novelty of our research, in fMRI acquisition and analysis, we added combined ERP/fMRI analyses to test for significant BOLD changes of each deviant stimulus (target, novel) with covariates defined using the P300 amplitude or latency. Given that both P300 amplitude and latency are informative to analyze the fMRI data, which suggest that measures of ERPs' amplitude may reflect the intensity of responses whereas measures of latency may reflect time duration of responses [20, 21], therefore, we used these two parameters as covariates to constrain fMRI data in the General Linear Model. In this way, we have found other forms of specific brain activation associated with P3a and P3b, such as hippocampal gyrus, bilateral fusiform gyrus, and insula, which extends prior studies.

The mainstream view is that selective attention is controlled by the combined brain area of the parietal and frontal lobes (e.g., SPL of the parietal lobe, FEP of the frontal eye area, SEP of the auxiliary eye area, etc.), which is perceived through top-down neural connections analysis [22, 23]. However, recent studies on time orientation have shown that attention control and attention regulation involve several other different neural mechanisms. Some researchers believe that the neural mechanisms of temporal and spatial expectations are independent. Spatial expectations affect early perception components, such as P1 and starvation time

expectations, which affect the late selection or response-related components, such as P300 components [24, 25]. Other researchers believe that the neural mechanisms of temporal and spatial expectations are partially independent and partially overlapping. They also admit that spatial expectations affect early perception components, but they believe that, according to different mission requirements, temporal expectations affect different stages of stimulus processing, including the early perception stage and late response selection stage [26, 28]. Coull and Nobre [27] used functional nuclear magnetic resonance imaging (fMRI) technology to find that there is a partial overlap between the neural mechanisms of spatial and temporal attention toward the task. Spatial and temporal attention showed hemispherical asymmetric activation on the right and left parietal lobe, respectively. And when paying attention to both time and space information, both parietal lobes are activated.

## 2. Methods

*2.1. Subjects.* Twenty-five healthy volunteers (average age =  $22 \pm 3$  years, range = 20–31 years, 22 men and 3 women), with no history of neurological or psychiatric diseases, participated in this study. All participants were carefully screened to ensure that they met the inclusion and exclusion criteria. They were all right-handed with normal or corrected to normal visual acuity. Finally, all participants provided written informed consent after the experimental procedures were fully explained.

*2.2. Study Procedure.* All participants completed the visual oddball task inside the scanner under a typical fMRI scanning environment with an MR-shielded EEG cap on. During this experiment, a visual three-stimulus oddball task was presented via the E-Prime software (Psychology Software Tools, Pittsburgh, USA). The novels were large colorful checkerboard squares (never repeated), and the targets were blue discs which were slightly larger than the standard blue discs. Within each task, a total of 128 stimuli were shown: 103 standard stimuli (85%), 15 target stimuli (12%), and 10 novel stimuli (8%). All stimuli were presented for 30 ms with an interstimulus interval of 1800 ms. All stimuli were in a pseudorandom sequence both within and between 25 subjects. Each participant laid inside the fMRI scanner with a button press box placed in the right hand, and they were instructed to press with their thumb when a larger blue disc was presented.

The radiology laboratory of Xijing Hospital uses a direct rear projection visual stimulation device. The projection system is completely nonmagnetic radio frequency shielding, which can be placed in the magnetic resonance scanning room and has an air-cooled heat dissipation device to ensure the stability of the projection. Compared with the general placement of projectors in the scanning control room, the rear projection method can avoid brightness attenuation and visual signal attenuation, maintain uniform brightness, ensure high-resolution visual signals, ensure the consistency of the experimental environment, and can easily be compared with others.

Magnetic resonance imaging brain functional audio visual stimulation system is mainly used in fMRI experiments to carry out visual and auditory stimulation. The system has good magnetic compatibility, can work normally in 3T MRI environment, and provides clear visual and auditory information to the subjects in the experimental process. According to the need, the experimenter can send the specific text, image, sound, and other experimental stimulus information materials to the subjects. At the same time, the feedback information of the subjects can be recorded in real time through key operation. There is also a clear voice communication between the experimenter and the subjects. The experimental stimulation task is synchronized with MRI scan, which ensures the time synchronization of brain function data and visual visual stimulation task. The system consists of visual stimulus host, data acquisition system, projection system, auditory stimulation module, stimulus feedback component, etc. (Figure 1).

The radiofrequency buttonless response control box designed according to ergonomics transmits the response of the subject directly to the computer through the CREATOR optical fiber transmission system. Event-related design can be used to statistically compare different types of cognitive tasks. CREATOR fiber optic synchronization system can send pulses to trigger the start of E-Prime stimulation program when the magnetic resonance scanning system starts scanning, so that the scanning and stimulation are strictly synchronized, providing a guarantee for accurate data analysis.

MRI examinations used the GE 3.0 T MRI scanner produced by the American General Medical Company and a 32-channel phased-array surface coil on the head. Conventional structural MRI examinations include T1W1 (TSE, TR/TE/TI 2510/9.6/1038 ms; flip angle 130°; layer thickness 3 mm; layer interval 0.3 mm; FOV 180 mm × 180 mm; matrix 256 × 256), T2WI (TSE, TR/TE 5000/93 ms, other scanning parameters are the same as before), and T2WI-FLAIR (TR/TE/TI = 9000/93/2500 ms; layer thickness 3.0 mm; layer interval 0.3 mm). In addition, in order to obtain a good gray-white contrast image, the scanning scheme also includes transversal IR sequence (TR/TE/TI = 1500/13/499 ms, layer thickness 4 mm, layer spacing 0.6 mm, flip angle 150°).

EEG data acquisition adopts 32-lead MR-compatible EEG system of German Brain Products company. The electrode is an antimagnetic material electrode. The electrode placement: the elastic electrode cap international 10–20 system is used to place the electrode. At the same time, the ECG is placed to record the heart rate change of the subject and identify electrocardiography traces. This study specifies that the scalp resistance is less than 10 k $\Omega$  and the acquisition frequency is 5000 Hz.

**2.3. EEG Acquisition and Analysis.** This laboratory uses a direct rear projection visual stimulation device. The projection system is completely nonmagnetic radio frequency shielding, which can be placed in the magnetic resonance scanning room, and there is an air-cooled heat dissipation device to ensure the stability of the projection. Compared with the general placement of projectors in the scanning control room, the rear projection method can avoid

brightness attenuation, visual signal attenuation, maintain uniform brightness, ensure high-resolution visual signals, ensure the consistency of the experimental environment, and can easily be compared with other The stimulation device is matched.

The CREATOR optical fiber synchronization system can send pulses to trigger the start of the E-Prime stimulation program when the magnetic resonance scanning system starts scanning, so that the scanning and stimulation are strictly synchronized, providing a guarantee for accurate data analysis.

The radiofrequency buttonless response control box designed according to ergonomics transmits the response of the subject directly to the computer through the CREATOR optical fiber transmission system. Event-related design can be used to statistically compare different types of cognitive tasks,

because it will generate a strong magnetic field when used in MRI work. According to the characteristics of magnetic resonance imaging (MRI), if the traditional VGA transmission cable is used, it will be interfered with by a large and strong magnetic field and cannot transmit signals normally. At the same time, the traditional VGA transmission cable will generate electrical and electromagnetic signals that will interfere with the normal operation of MRI. Therefore, CREATOR fiber transmission has become a wise choice. The biggest advantage of optical fiber transmission is the following. (1) It solves the long-distance or ultralong distance (tens of kilometers) VGA signal transmission and is the lossless transmission. (2) In many specific occasions, such as the electromagnetic environment, optical fiber transmission is very bad. The antijamming feature of the system can be brought into full play, expanding the range of VGA signal transmission applications. (3) Compared with the traditional analog cable mode, the thinking mode of all-digital and all-fibers is self-evident, and it will certainly bring a lot of new opportunities and growth points, which has been confirmed. (4) On many occasions, the comprehensive cost of digital fiber mode can be compared with analog and cable modes and may even be lower, which provides the possibility for a large number of future applications.

VGA signal transmission system uses CREATORVGA-F500 T/RVGA optical fiber transceiver, computer output, uses its VGA port to lead analog signals to VGA-F500TVGA optical fiber transceiver, and converts VGA signal to optical signal transmission. VGA-F500 R is responsible for optical revert to VGA signal. The signal quality after digital transmission can be greatly improved, and problems such as smearing, blurring, and ghosting, which were common in analog transmission in the past, will not occur.

Electroencephalogram (EEG) data was recorded simultaneously with fMRI data using a BrainAmp MR-compatible EEG system (BrainProducts GmbH, Munich, Germany), which includes a 32-channel MR-compatible amplifier and an MR-compatible EEG cap. The EEG cap included 30 scalp electrodes, one electrooculogram (EOG), and one electrocardiogram (ECG). The EOG electrode was placed beneath the subject's left eye to monitor the eye blink, and the ECG electrode was placed over the subject's back. The common recording reference was referenced online to



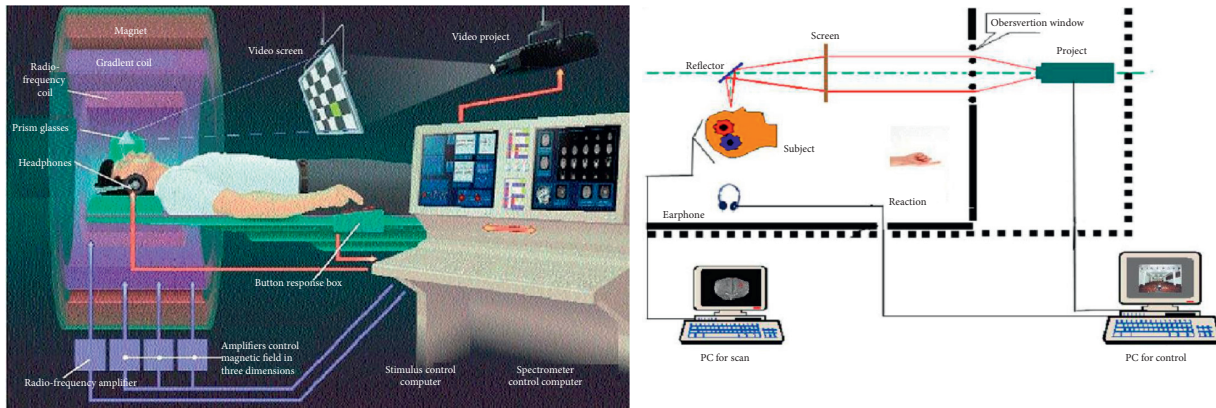


FIGURE 1: Study procedure.

the frontocentral midline electrode. Impedance for the electrodes was kept below 5 k $\Omega$  by using the electrode paste. The EEG signals were digitally sampled at 5000 Hz.

Next, the EEG data were analyzed by BrainVision Analyzer software (Brain Products GmbH, Gilching, Germany). First, scanner gradient artifacts and pulse-related artifacts were removed using an adaptive average subtraction method [20] and an algorithm based on a constrained independent component analysis (ICA) method [21]; meanwhile, the EEG signals were down-sampled at 250 Hz. Second, EEG signals, rereferenced to right and left mastoids (M1 M2), were low-pass filtered (between 0.01 and 35 Hz). Finally, eye blinks were manually detected and then regressed from the EEG signal.

The ERPs were computed for the target, novel, and standard stimuli, respectively. The P3a (novel effect) and P3b (target effect) components were separately defined as the largest positive deflection during the time window between -200 and 800 msec. The grand average ERP waveforms for the target and novel stimuli at typical three midline electrodes were all used to acquire our results in the study.

**2.4. fMRI Acquisition and Analysis.** fMRI data were acquired by a GE 3T MR scanner (GE Health Care, Milwaukee, WI) at the Department of Radiology of Xijing Hospital, Xi'an, China. The functional images were collected using a gradient-echo planar imaging sequence with the following parameters: TR = 2000 ms, TE = 30 ms, FOV = 240 mm  $\times$  240 mm, data matrix = 64  $\times$  64, flip angle = 90', slices = 45, and 150 volumes. The anatomical images were also obtained (TR/TE: 8.2 ms/3.18 ms, field of view: 256  $\times$  256 mm<sup>2</sup>, matrix size: 512  $\times$  512, flip angle = 9°, in-plane resolution: 0.5  $\times$  0.5 mm<sup>2</sup>, slice thickness = 1 mm, 196 sagittal slices).

Preprocessing and data analysis were both performed with the SPM8 software (Wellcome Department of Cognitive Neurology, London) implemented in MATLAB. First, in order to ensure the stability of the radial magnetic field, we discard the first four volumes of whole-brain images. Second, images were corrected for the acquisition delay between the slices and were realigned to the first volume. After the adjustment for movement-related effects, functional and

anatomical data were coregistered, spatially normalized into the standard stereotactic MNI space, and spatially smoothed with a 6 mm full width half maximum (FWHM) Gaussian kernel.

The event-related design (e-fMRI) was reported to be highly sensitive to the oddball task-related. A conventional analysis was conducted to assess the related activation in BOLD response associated with the processing of three stimuli. Each stimulus type (target, standard, and novel) was used as a regressor in the GLM after convolution with the hemodynamic response function. And the six movement parameters were used as regressors of noninterest. Target and novel functional activation was discriminated from baseline activation through the regression analysis. Regions that are significantly more activated by the targets relative to the novels were also given using a fixed-effect analysis ( $p < 0.001$  uncorrected and the cluster size  $> 50$  voxels).

### 3. Results

**3.1. Behavioral Data.** Based on the statistical data, for the target stimuli in the task, all subjects responded correctly to 97.6% (SD: 5.7%) with a response time of 408 ms (SD: 48 ms), which indicated that they were able to concentrate on their tasks during the EEG recording and MRI scans.

**3.2. ERP Data.** Figure 2 shows the ERP results computed across 25 subjects for three electrodes (Fz, Cz, and Pz) and three stimuli (standard, target, and novel). The figure illustrates that P300 ERPs could be detected at the frontocentral (Fz), central (Cz), and centroparietal (Pz) electrodes for targets and novels. Using paired  $t$ -test, the novel-related P300 ERP (P3a) had a higher amplitude and decreased more slowly than the target-related P300 ERP (P3b), at Cz and Fz (all  $p < 0.01$ ), but for Pz, the P3b ERP had a significantly higher amplitude ( $p < 0.01$ ). P3a arose earlier than P3b at Cz and Fz (all  $p < 0.01$ ), but the opposite phenomenon was observed for Pz ( $p < 0.01$ ). Figure 2 demonstrates that P3a and P3b have distinct topographic amplitude distributions. The P3b activated frontal areas, while P3a activated frontoparietal areas. These topographic variations may be induced by different functional areas.

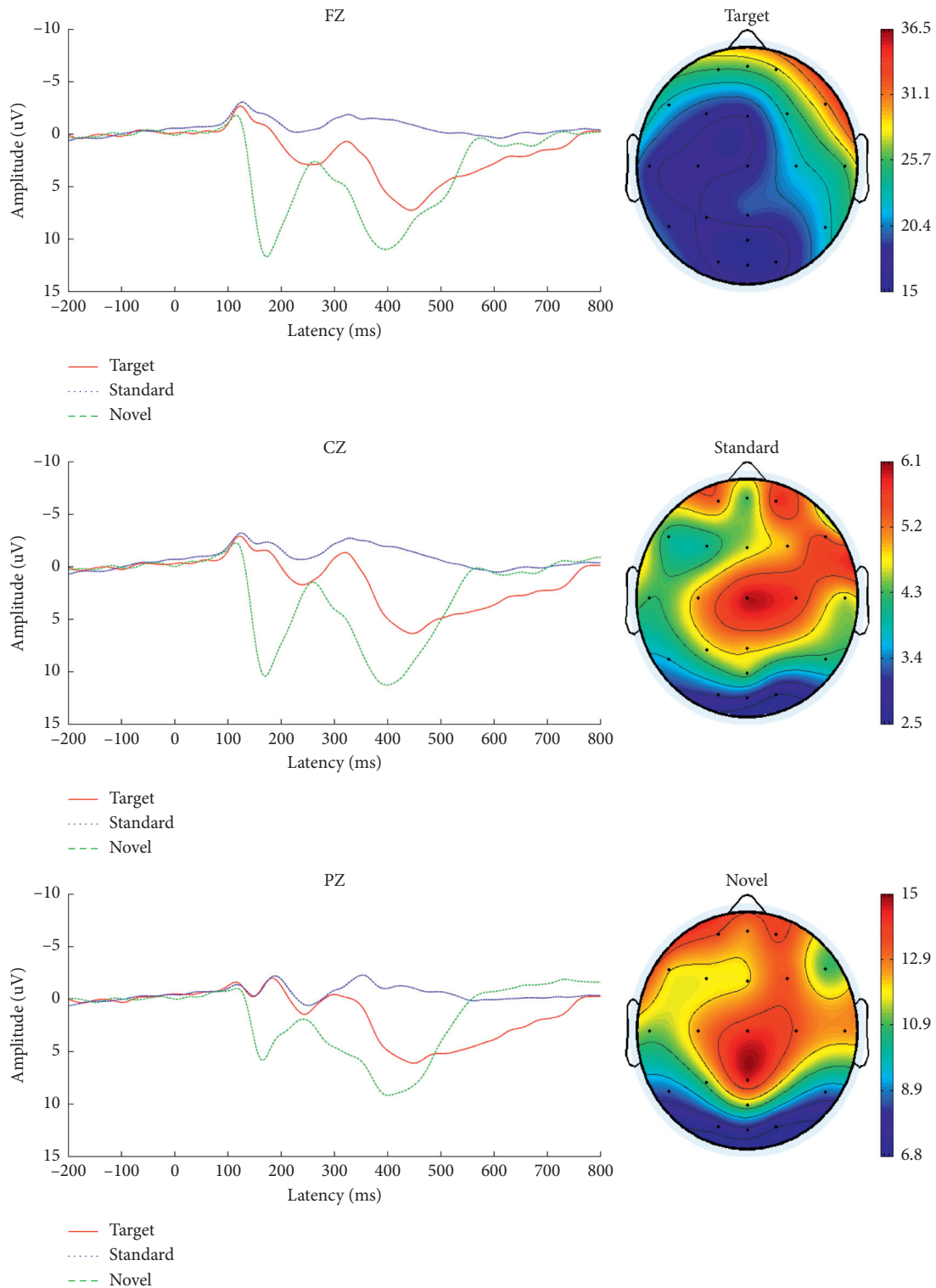


FIGURE 2: (a) Grand-averaged waveforms (mean across the 25 participants and 3 types of stimuli) on Pz, Cz, and Fz electrodes: target stimuli (red), standard (blue), and novels (green). (b) Topography distributions for the mean target (A), standard (B), and novel (C) amplitude.

3.3. *fMRI Data.* Figure 3(a) and Table 1 illustrate the activation for the novel stimuli. The novels activated the left frontoparietal areas, left precuneus areas, and the right insula. Figure 3(b) and Table 2 illustrate the activation for the target stimuli. The targets activated the bilateral fusiform gyrus, left frontal areas, and

bilateral insula. Figure 3(c) and Table 3 illustrate the activation for the novel-target contrast. These included the left frontoparietal areas, the left postcentral, and the left fusiform.

Gray reflects the background baseline state of the cerebral cortex; the darker the color, the closer to orange-red,

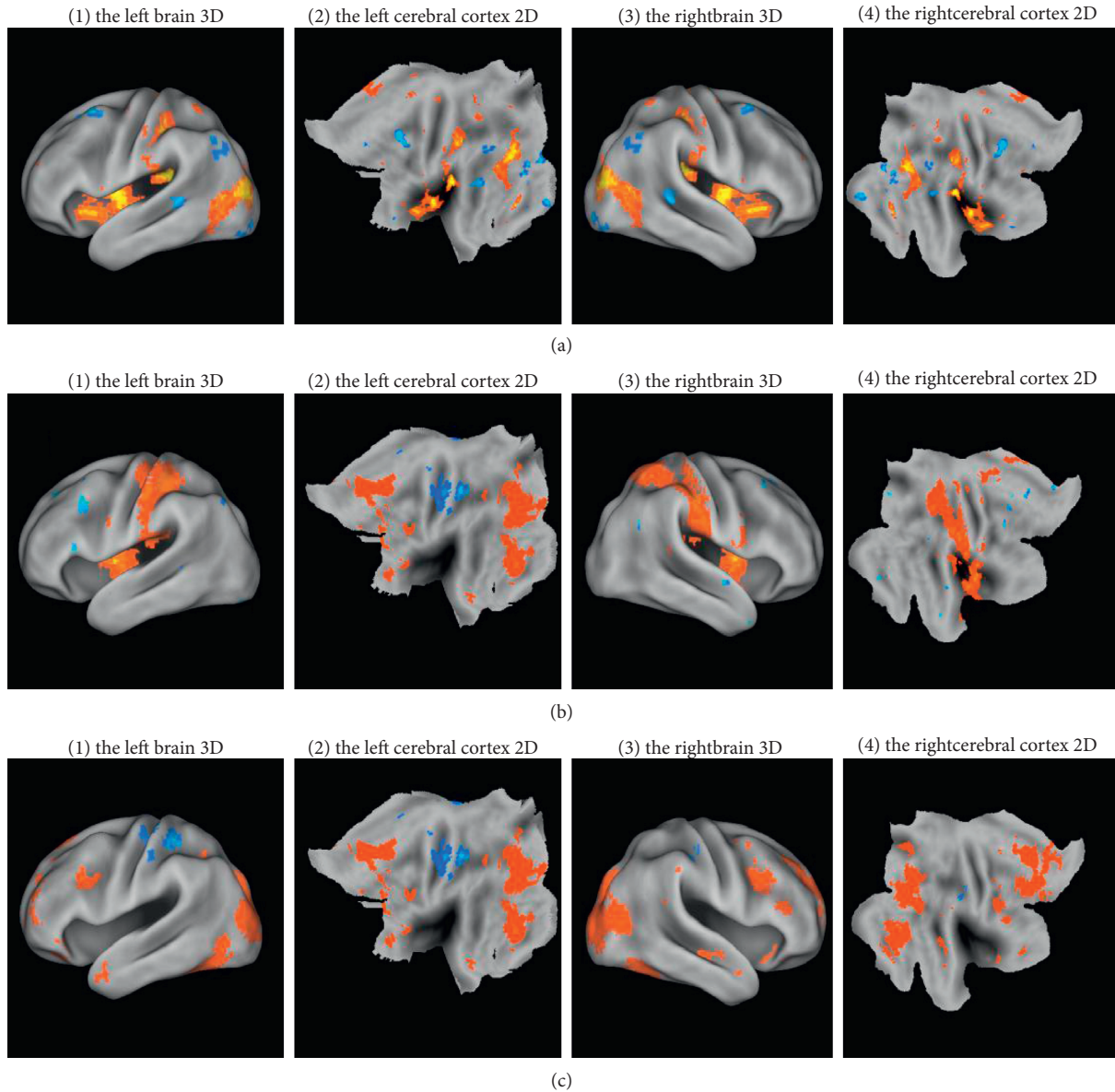


FIGURE 3: Functional activation 3D maps of left and right brain obtained for the novel (a), target (b), and target-novel contrast (c). (a) Brain regions showing activity to the novel stimuli. (b) Brain regions showing activity to the target stimuli. (c) Brain regions showing activity to the novel-target contrast. Reported activation became all thresholds at  $p < 0.001$  uncorrected. Gray reflects the background baseline state of the cerebral cortex; the darker the color, the closer to orange-red, indicating that the increased rate of blood oxygen concentration in the brain area is more significant, indicating that the area is being activated. The lighter the color, the closer to blue, which means that the rate of decrease in blood oxygen concentration in this brain area is more significant, indicating that this area is negatively activated. a1: in the new heterostimulus task state, the activation area of the left brain 3D model; a2: in the new heterostimulus task state, the activation area of the 2D model of the left cerebral cortex; a3: in the new heterostimulus task state, the activation area of the right brain 3D model; a4: in the new heterostimulus task state, the activation area of the 2D model of the right cerebral cortex; b1: in the target stimulation task state, the activation area of the left brain 3D model; b2: in the target stimulation task state, the activation area of the 2D model of the left cerebral cortex expanded; b3: in the target stimulation task state, the activation area of the right brain 3D model; b4: in the target stimulation task state, the activation area of the 2D model of the right cerebral cortex; c1: for fMRI image data, using target stimulation-new stimulation, the activation area of the left brain 3D model; c2: for fMRI image data, using target stimulation-differential stimulation, the activation area of the 2D model of the left cerebral cortex; c3: for fMRI image data, use target stimulation-new stimulation, the activation area of the right brain 3D model; c4: for fMRI image data, using target stimulation-differential stimulation, the activation area of the 2D model of the right cerebral cortex.

TABLE 1: Regions of P300 FMRI activation to the novels.

| Peak MNI coordinate | Anatomical area     | $K$ (cluster extent) | Peak $Z$ score |
|---------------------|---------------------|----------------------|----------------|
| -27 18 51           | Frontal_Mid_L (aal) | 107                  | 5.0358         |
| 0 -60 36            | Precuneus_L (aal)   | 69                   | 4.1338         |
| -39 -3 12           | Insula_R (aal)      | 53                   | 4.254          |
| -51 -36 39          | Parietal_Inf_L(aal) | 57                   | 4.0481         |

The coordinates  $x$ ,  $y$ , and  $z$  (mm) are all given in the Montreal Neurological Institute (MNI) standard stereotactic space. The results are all significant at the voxel level  $p < 0.001$  uncorrected and for the least cluster extent of 50 voxels. Overall  $z$  score:  $z = (X-\mu)/\sigma$  where  $X-\mu$  is the deviation from the mean and  $\sigma$  represents the standard deviation. Sample  $z$  score.

TABLE 2: Regions of P300 FMRI activation to the targets.

| Peak MNI coordinate | Anatomical area         | $K$ (cluster extent) | Peak $Z$ score |
|---------------------|-------------------------|----------------------|----------------|
| -27 -75 -9          | Fusiform_L (aal)        | 162                  | 4.6535         |
| -48 21 21           | Frontal_Inf_Tri_L (aal) | 101                  | 4.0602         |
| -27 -27 54          | Precentral_L (aal)      | 92                   | 3.85           |
| -36 12 54           | Frontal_Mid_L (aal)     | 107                  | 4.3317         |
| 14 -80 -8           | Fusiform_R (aal)        | 147                  | 4.7589         |
| 32 8 14             | Insula_R (aal)          | 63                   | 3.7151         |
| -34 0 12            | Insula_L (aal)          | 54                   | 3.5945         |

The coordinates  $x$ ,  $y$ , and  $z$  (mm) are all given in the Montreal Neurological Institute (MNI) standard stereotactic space. The results are all significant at the voxel level  $p < 0.001$  uncorrected and for the least cluster extent of 50 voxels.

TABLE 3: Regions of P300 FMRI activation to novels-targets contrast.

| Peak MNI coordinate | Anatomical area         | $K$ (cluster extent) | Peak $Z$ score |
|---------------------|-------------------------|----------------------|----------------|
| -33 -48 -18         | Fusiform_L (aal)        | 92                   | 4.0206         |
| -27 -45 72          | Postcentral_L (aal)     | 212                  | 4.1919         |
| -24 -60 48          | Parietal_Sup_L (aal)    | 69                   | 3.5715         |
| -51 33 21           | Frontal_Inf_Tri_L (aal) | 72                   | 4.18           |

The coordinates  $x$ ,  $y$ , and  $z$  (mm) are all given in the Montreal Neurological Institute (MNI) standard stereotactic space. The results are all significant at the voxel level  $p < 0.001$  uncorrected and for the least cluster extent of 50 voxels.

indicating that the increased rate of blood oxygen concentration in the brain area is more significant, indicating that the area is being activated. The lighter the color, the closer to blue, which means that the rate of decrease in blood oxygen concentration in this brain area is more significant, indicating that this area is negatively activated.

#### 4. Discussion

The current study used the simultaneous ERP/fMRI data collected in a visual three-stimulus oddball task to examine the differences of spatiotemporal neural dynamics between the P3a and P3b and the neuropsychological origins underlying these two ERPs.

The results obtained from the electrical recording, as shown in Figure 2, showed that P3a and P3b have different topographic amplitude distributions, which were in line with previous studies [29, 30]. As expected, two P300 ERPs were observed at middle sites (Fz, Cz, Pz, C3, and C4) for the novel and target conditions, and statistical differences in amplitudes between targets and novels were found [30, 31]. P3a and P3b data, using a difficult target/standard discrimination task in a three-stimulus paradigm from 120 healthy young adults, have shown that P3a has a central maximum, whereas P3b has a parietal maximum, and the

peak latency for both was shorter over the frontal and longer over the parietal sites [7, 19, 32].

Simultaneous fMRI and ERP findings demonstrated different brain regions related to different stimuli. P3a and P3b generation is assumed to stem from frontal and temporal/parietal activation. In the present study, P3a was related to the frontal lobe. It is proposed that P3a may be generated when rare or novel stimuli are processed if sufficient attentional focus is engaged (see Figure 3 and Tables 1–3) [33]. Many studies suggest that a frontal attention mechanism governs neural responsivity to novels [34] and then implies top-down control [35]. Attentional resources used to maintain memory items in parietal regions may result from response organization produced by bottom-up processing [7]. P3b was associated with parietal regions, prefrontal cortex, and medial temporal regions [36]. Evidence suggested that P3b appears when subsequent attentional resource activation promotes memory operations in temporal-parietal areas [37, 38]. Thus, the simultaneous ERP/fMRI findings extended the proposed model which suggested that P300 comprises an early attention process stemming from a frontal working memory representational change to produce the P3a, and the attention-driven stimulus signal is then transmitted to temporal and parietal structures, which were related to P3b [36, 39–41].

The limitations of this study should also be mentioned. First, 19 of the 25 subjects, from the Fourth Military Medical University, showed higher scores ( $128 \pm 5$ ) in the intelligence tests, which may have affected the result. The subsequent research can explore the potential impact of intelligence on ERP. Second, ERP analysis was restricted to only 3 electrodes without taking into account the full amount of the temporospatial information included in the ERP data. Third, gender has been shown to modulate the P300 component in emotional oddball tasks [42], but our sample included both female and male participants. Consequently, it is unknown whether our results were affected by the gender effect.

Furthermore, in order to better promote the fusion research of EEG-fMRI, there are still some deficiencies in the research of this article; the specific improvements are as follows.

Offline antialiasing processing is performed on the three artifact components of pulse artifact, gradient artifact, and electrocardiographic artifact. Due to the tedious experiment of synchronously collecting EEG signals in the magnetic resonance scanner, the limb movement or Mental state, etc. may introduce many unknown artifacts. Therefore, how to simplify the collection steps and minimize the generation of unknown artifacts is a problem to be improved in the future.

The amount of experimental calculation data of the synchronous EEG-fMRI brain-computer interface system in this article is large. In order to have a better real-time effect and reduce the amount of calculation, we only conducted the lead of the specific brain area where the P300 signal is active and analyze and ignore the information that other leads may reflect. Therefore, this topic should continue to improve the research. In addition, only 32-lead amplifiers were used in this experiment. In order to observe EEG signals more accurately, it may be considered to introduce 64-lead or 128-lead amplifiers for acquisition to obtain more detailed EEG information.

Because this experiment is designed for the P300 signal, but the research methods are universal, the next step should consider applying the synchronous EEG-fMRI brain-computer interface system to more experimental paradigms, such as emotion recognition and SSVEP.

Since this article only focuses on the EEG signals in the magnetic resonance scanning environment, most of the introduced artifacts are eliminated through real-time artifact removal processing. Therefore, in future research, the fMRI signal should be processed and the electroencephalogram signal and the fMRI signal should be fused to develop a brain signal that can observe multimodalities in real time, thus laying a foundation for humans to understand the brain from multiple angles and promote the development of brain science.

## 5. Conclusion

The integration of event-related potential (ERP) and functional magnetic resonance imaging (fMRI) can be used as a bridge between the well-established field of evoked cognitive potentials and the fast-growing field of fMRI. The simultaneous EEG/fMRI data provided further insight into

processes underlying the functional brain activation. In the oddball task, our results showed that P300 ERP induced by target stimulation and novel stimulation can be detected at the Fz, Cz, and Pz electrodes. At Cz and Fz, compared with P3b related to the target stimulus, the P3a related to the novel stimulus has a higher amplitude and the waveform declines more slowly. But at Pz, P3b has a higher amplitude than P3a. P3a appeared earlier than P3b at Cz and Fz, but the opposite phenomenon was observed at the Pz electrode. The activated brain regions of P3a included the left frontal-parietal lobe region, left anterior wedge lobe region, and right insula, while the target-driven P3b was significantly associated with BOLD changes in the bilateral fusiform gyrus, the left frontal region, and the bilateral insula. In summary, the current study replicates and extends prior studies.

## Abbreviations

|           |                                       |
|-----------|---------------------------------------|
| EEG:      | Electroencephalogram                  |
| EEG-fMRI: | Simultaneous EEG and fMRI recordings  |
| MRI:      | Magnetic resonance imaging            |
| BOLD:     | Blood oxygen level-dependent          |
| fMRI:     | Functional magnetic resonance imaging |
| ERP:      | Event-related potential               |
| MNE:      | Minimum norm estimate                 |
| ECD:      | Equivalent current dipole             |
| DCD:      | Distributed current density           |
| HRF:      | Hemodynamic response function.        |

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

The authors are grateful to Professor Wei Qin for reading and correcting the manuscript and for fruitful criticisms. The authors thank Professor Junye Liu who designed and participated in the experiment in this study. Special thanks are due to Rui Zhao for her help in processing and interpreting the EEG data and Dian Dai and Xinxin Zhang for their help in the acquisition of fMRI data. This work was supported by the Department of Radiology of Xijing Hospital that provided the MRI device. This project was supported by the National Natural Science Foundation of China (Grant no. 61806210) and the National Social Science Foundation of China (Grant no. 15BSH023).

## References

- [1] J. Polich, "Updating P300: an integrative theory of P3a and P3b," *Clinical Neurophysiology*, vol. 118, no. 10, pp. 2128–2148, 2007.
- [2] E. Donchin, "Surprise!? Surprise?" *Psychophysiology*, vol. 18, no. 5, pp. 493–513, 1981.

- [3] S. Adachi, K. Morikawa, and H. Nittono, "Event-related potentials elicited by unexpected visual stimuli after voluntary actions," *International Journal of Psychophysiology*, vol. 66, no. 3, pp. 238–243, 2007.
- [4] R. Johnson, "On the neural generators of the P300 component of the event-related potential," *Psychophysiology*, vol. 30, no. 1, pp. 90–97, 1993.
- [5] E. Snyder and S. A. Hillyard, "Long-latency evoked potentials to irrelevant, deviant stimuli," *Behavioral Biology*, vol. 16, no. 3, pp. 319–331, 1976.
- [6] N. K. Squires, K. C. Squires, and S. A. Hillyard, "Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man," *Electroencephalography and Clinical Neurophysiology*, vol. 38, no. 4, pp. 387–401, 1975.
- [7] M. A. Conroy and J. Polich, "Normative variation of P3a and P3b from a large sample," *Journal of Psychophysiology*, vol. 21, no. 1, pp. 22–32, 2007.
- [8] H. Gaeta, D. Friedman, and G. Hunt, "Stimulus characteristics and task category dissociate the anterior and posterior aspects of the novelty P3," *Psychophysiology*, vol. 40, no. 2, pp. 198–208, 2003.
- [9] E. Yago, C. Escera, K. Alho, M. H. Giard, and J. M. Serra-Grabulosa, "Spatiotemporal dynamics of the auditory novelty-P3 event-related brain potential," *Cognitive Brain Research*, vol. 16, no. 3, pp. 383–390, 2003.
- [10] V. P. Clark, S. Fannon, S. Lai, and R. Benson, "Paradigm-dependent modulation of event-related fMRI activity evoked by the oddball task," *Human Brain Mapping*, vol. 14, no. 2, pp. 116–127, 2001.
- [11] J. R. Foucher, H. Otzenberger, and D. Gounot, "The BOLD response and the gamma oscillations respond differently than evoked potentials: an interleaved EEG-fMRI study," *BMC Neuroscience*, vol. 4, no. 1, p. 22, 2003.
- [12] J. R. Foucher, H. Otzenberger, and D. Gounot, "Where arousal meets attention: a simultaneous fMRI and EEG recording study," *Neuroimage*, vol. 22, no. 2, pp. 688–697, 2004.
- [13] K. A. Kiehl, K. R. Laurens, T. L. Duty, B. B. Forster, and P. F. Liddle, "Neural sources involved in auditory target detection and novelty processing: an event-related fMRI study," *Psychophysiology*, vol. 38, no. 1, pp. 133–142, 2001.
- [14] R. T. Knight, M. F. Grabowecy, and D. Scabini, "Role of human prefrontal cortex in attention control," *Advances in Neurology*, vol. 66, pp. 21–26, 1995.
- [15] Y. Li and W. Wang, "Source analysis of P3a and P3b components to investigate interaction of depression and anxiety in attentional systems," *Scientific Reports*, vol. 5, no. 1, Article ID 17138, 2015.
- [16] R. Verleger, P. Jas'kowskis, and E. Wascher, "Evidence for an integrative role of P3b in linking reaction to perception," *Journal of Psychophysiology*, vol. 19, p. 150, 2005.
- [17] S. Yamaguchi and R. T. Knight, "Effects of temporal-parietal lesions on the somatosensory P3 to lower limb stimulation," *Electroencephalography and Clinical Neurophysiology/ Evoked Potentials Section*, vol. 84, no. 2, pp. 139–148, 1992.
- [18] J. Polich and M. D. Comerchero, "P3a from visual stimuli: typicality, task, and topography," *Brain Topography*, vol. 15, no. 3, pp. 141–152, 2003.
- [19] M. Soltani and R. T. Knight, "Neural origins of the P300," *Critical Reviews in Neurobiology*, vol. 14, pp. 199–224, 2000.
- [20] P. J. Allen, O. Josephs, and R. Turner, "A method for removing imaging artifact from continuous EEG recorded during functional MRI," *NeuroImage*, vol. 12, no. 2, pp. 230–239, 2000.
- [21] Y. Leclercq, J. Schrouff, Q. Noirhomme, P. Maquet, and C. Philipps, "fMRI artefact rejection and sleep scoring toolbox," *Computational Intelligence and Neuroscience*, vol. 2011, p. 11, Article ID 598206, 2011.
- [22] A. Gazzaley and A. C. Nobre, "Top-down modulation: bridging selective attention and working memory," *Trends in Cognitive Sciences*, vol. 16, no. 2, pp. 129–135, 2012.
- [23] S. Kastner, P. de Weerd, and L. G. Ungerleider, "Texture segregation in the human visual cortex: a functional MRI study," *Journal of Neurophysiology*, vol. 83, no. 4, pp. 2453–2457, 2000.
- [24] J. R. Doherty, A. Rao, M. M. Mesulam, and A. C. Nobre, "Synergistic effect of combined temporal and spatial expectations on visual attention," *The Journal of neuroscience: the official journal of the Society for Neuroscience*, vol. 25, no. 36, pp. 8259–8266, 2005.
- [25] I. C. Griffin, C. Miniussi, and A. C. Nobre, "Multiple mechanisms of selective attention: differential modulation of stimulus processing by attention to space or time," *Neuropsychologia*, vol. 40, no. 13, pp. 2325–2340, 2002.
- [26] A. Correa, J. Lupiáñez, E. Madrid, and P. Tudela, "Temporal attention enhances early visual processing: a review and new evidence from event-related potentials," *Brain Research*, vol. 1076, no. 1, pp. 116–128, 2006.
- [27] J. T. Coull and A. C. Nobre, "Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI," *Journal of Neuroscience*, vol. 18, no. 18, pp. 7426–7435, 1998.
- [28] J. Rimmele, E. Sussman, and H. Jolsvai, "Auditory target detection is affected by implicit temporal and spatial expectations," *Journal of Cognitive Neuroscience*, vol. 23, no. 5, pp. 1136–1147, 2010.
- [29] J. Polich, "P300 in clinical applications," in *Electroencephalography: Basic Principles, Clinical Applications and Related Fields*, E. Niedermeyer and F. Lopes Da Silva, Eds., pp. 1073–1091, Williams & Wilkins, Baltimore, MD, USA, 1999.
- [30] H. Otzenberger, D. Gounot, and J. R. Foucher, "P300 recordings during event-related fMRI: a feasibility study," *Cognitive Brain Research*, vol. 23, no. 2–3, pp. 306–315, 2005.
- [31] T. Eichele, K. Specht, M. Moosmann et al., "Assessing the spatiotemporal evolution of neuronal activation with single-trial event-related potentials and functional MRI," *Proceedings of the National Academy of Sciences*, vol. 102, no. 49, pp. 17798–17803, 2005.
- [32] D. Linden, "The P300: where in the brain is it produced and what does it tell us?" *Neuroscientist*, vol. 5, no. 6, pp. 563–76, 2003.
- [33] M. Brázdil, R. Roman, P. Daniel, I. Rektor, and I. Rektor, "Intracerebral somatosensory event-related potentials: effect of response type (button pressing versus mental counting) on P3-like potentials within the human brain," *Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology*, vol. 114, no. 8, pp. 1489–96, 2003.
- [34] S. Suwazono, L. Machado, and R. T. Knight, "Predictive value of novel stimuli modifies visual event-related potentials and behavior," *Clinical Neurophysiology*, vol. 111, no. 1, pp. 29–39, 2000.
- [35] C. Bledowski, D. Prvulovic, R. Goebel, F. E. Zanella, and D. E. J. Linden, "Attentional systems in target and distractor processing: a combined ERP and fMRI study," *Neuroimage*, vol. 22, no. 2, pp. 530–540, 2004.

- [36] J. K. Wynn, A. M. Jimenez, B. J. Roach et al., "Impaired target detection in schizophrenia and the ventral attentional network: findings from a joint event-related potential-functional MRI analysis," *NeuroImage: Clinical*, vol. 9, pp. 95–102, 2015.
- [37] M. Brazdil, I. Rektor, P. Daniel, M. Dufek, and P. Jurak, "Intracerebral event-related potentials to subthreshold target stimuli," *Clin Neurophysiol*, vol. 112, pp. 650–61, 2001.
- [38] L. Squire and E. Kandel, *Memory From Mind to Molecules*, Scientific American Library, New York, NY, USA, 1999.
- [39] K. A. Kiehl, M. C. Stevens, K. R. Laurens, G. Pearlson, V. D. Calhoun, and P. F. Liddle, "An adaptive reflexive processing model of neurocognitive function: supporting evidence from a large scale (n=100) fMRI study of an auditory oddball task," *Neuroimage*, vol. 25, no. 3, pp. 899–915, 2005.
- [40] S. Campanella, M. Bourguignon, P. Peigneux et al., "BOLD response to deviant face detection informed by P300 event-related potential parameters: A simultaneous ERP-fMRI study," *NeuroImage*, vol. 71, pp. 92–103, 2013.
- [41] S. G. Horowitz, "Correlations and dissociations between BOLD signal and P300 amplitude in an auditory oddball task: a parametric approach to combining fMRI and ERP," *Magnetic Resonance Imaging*, vol. 20, pp. 319–325, 2001.
- [42] S. Campanella, L. Falbo, M. Rossignol et al., "Sex differences on emotional processing are modulated by subclinical levels of alexithymia and depression: A preliminary assessment using event-related potentials," *Psychiatry Research*, vol. 197, no. 1-2, pp. 145–153, 2012.

## Research Article

# Intelligent Digital Currency and Dynamic Coding Service System Based on Internet of Things Technology

Shanshen Li <sup>1</sup> and Xin Jing <sup>2</sup>

<sup>1</sup>School of Economics and Management, Xi'an Shiyou University, Xi'an 710065, Shaanxi, China

<sup>2</sup>Law School, Shanghai University of Finance and Economics, Shanghai 200433, China

Correspondence should be addressed to Xin Jing; [jingxin@163.sufe.edu.cn](mailto:jingxin@163.sufe.edu.cn)

Received 22 October 2020; Revised 28 November 2020; Accepted 30 November 2020; Published 22 December 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Shanshen Li and Xin Jing. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The amazing rise of digital currency is not only favored by investors but also attractive to lawbreakers for its anonymity and decentralization. This paper mainly discusses the intelligent digital currency and dynamic coding service system based on Internet of Things technology. In this paper, the RDCAR algorithm is used to realize the routing discovery process of the wireless network. When the intermediate node receives the RREQ message, first of all, to avoid the loop, it checks whether the same RREQ message has been introduced. If it has received it, it will discard it. Otherwise, it will cache the message and attach its own neighbor node list to the signal-to-noise ratio of the channel link, update the RREQ message, and broadcast it. The payment cipher is managed by the bank. When the user opens an account, the bank registers and sends it to the user. The key is generated by the algorithm chip, and the public key is kept in the bank background server. When the bill is delivered to the bank, the bank inputs all the elements on the bill on the counter terminal and transmits it to the verification machine for verification through the bank network. If the verification is correct, it indicates that the bill is indeed issued by the customer, and all bill elements are correct, and payment can be made. The node operation protocol of public chain and alliance chain maintains the operation of the Internet of Things system. The nodes of alliance chain generate new blocks according to the interval of 30 s. When the node fails to complete the block generation within 30 s, it will rotate to the next node. The mkfile command is used to generate 16b, 1 KB, 1 MB, and 1 GB files as input. The peak speed of the encoding service system is about 370 mb/s. The results show that the system designed in this study is robust and suitable for complex trading environment.

## 1. Introduction

With the rise of the wave of digital currency in recent years, some underlying technologies related to it, such as blockchain technology and distributed accounting methods, also show broad application prospects. Digital currency is moving from theory to reality, and its feasibility and security are being tested. As a typical application of the Internet of Things in the financial economy, the Bitcoin system has attracted much attention due to its decentralized, open, and transparent characteristics. Various cryptocurrencies such as Monero coins and dark coins emerge in endlessly. In addition, central banks and commercial banks in various countries have also used the underlying advanced technology of Bitcoin as a reference, planning to carry out

research on legal digital currency, so as to improve the national digital financial system. This shows that the future development of digital currency has broad prospects, but the accompanying digital currency security and privacy issues have become increasingly prominent.

The amazing growth of encrypted digital currency is not only favored by investors but its anonymity and decentralization are also quite attractive to criminals. Although some countries have considered introducing digital currency regulatory policies, the attitudes of different countries are different. Facing the diversified needs of people's life and work, research and deployment of legal digital currency are urgently needed. However, unlike Bitcoin's non-real-time, lightweight transaction information, small transaction volume, and low sensitivity, legal digital currency is circulated



throughout the country and the world, and its security and privacy issues are more prominent, and legal digital currency and its related applications should be suitable for more complex international environments. Therefore, how to strengthen the robustness of the legal digital currency system and build a harmonious credit society by learning from the security and privacy protection measures of the Bitcoin system and Bitcoin wallet is the research direction of our future work. Digital currency is backed by national credit, which can synchronize online and offline to the greatest extent, and maximizes the convenience and security of transactions.

Utilizing the advantages of distributed architecture and proximity to end users, fog/edge computing can provide faster response and higher quality of service for IoT applications. Lin et al. believed that the Internet of Things based on fog/edge computing will become the future infrastructure for the development of the Internet of Things. In order to develop the IoT infrastructure based on fog/edge computing, they first studied the architecture related to the IoT, supporting technologies and issues, and then explored the integration of fog/edge computing and the IoT. They gave a comprehensive overview of the Internet of Things in terms of system architecture, supporting technology, security, and privacy issues and studied the integration of fog/edge computing with the Internet of Things and applications. Their research lacks data [1]. In order to overcome the scalability problem of the traditional IoT architecture, Sun et al. proposed a novel method for mobile edge computing. At the same time, they proposed a layered fog computing architecture in each fog node to provide flexible IoT services while maintaining user privacy; each user's IoT device is associated with the proxy VM (located in the fog node), and proxy VM collects, classifies, and analyzes the raw data stream of the device, converts it into metadata, and then transmits the metadata to the corresponding application VM (owned by the IoT service provider). Each application VM receives corresponding metadata from a different proxy VM. Their research lacks practice [2]. Yang et al. believed that the Internet of Things (IoT) is ubiquitous in our daily lives. In order to protect the security of IoT devices, they have conducted a lot of research work to deal with these problems and find better ways to eliminate these risks, or at least minimize its impact on user privacy and security requirements. Their investigation consists of four parts. The most relevant limitations of IoT devices and their solutions will be discussed first. Then, the classification of IoT attacks will be introduced. Secondly, it will focus on the mechanism and architecture of authentication and access control. Finally, the security issues in different layers will be analyzed. Their research process is too complicated [3]. Yaqoob et al. discussed the architecture of the Internet of Things. In this case, first of all, they investigate, focus on, and report on the recent major research progress in the IoT architecture and then classify the IoT architecture and design a taxonomy based on important parameters (such as applications, supporting technologies, business goals, architecture requirements, network topology, and IoT platform architecture types).

They identified and outlined the key requirements of the future IoT architecture and discovered and introduced some outstanding case studies on the Internet of Things. Finally, they listed and outlined the future research challenges. Their research has no practical significance [4].

This research mainly discusses the intelligent digital currency and dynamic coding service system based on the Internet of Things technology. This research is mainly based on the RDCAR algorithm to realize the route discovery process of the wireless network. When the intermediate node receives the RREQ message, first, to avoid loops, check whether the same RREQ message has been introduced. If it is received, discard it. Otherwise, buffer the message and append its own neighbor node list, corresponding to the signal-to-noise ratio of the channel link, and update RREQ message and broadcast it. The payment cipher is managed by the bank. When the user opens an account, the bank registers and sends it to the user, and the key is generated by the algorithm chip, and the public key is stored in the bank's back-end server. When the bill is delivered to the bank, the bank enters the various elements of the bill on the counter terminal and transmits it to the verification machine through the bank network for verification. If the verification is correct, it indicates that the bill is indeed issued by the customer, and the bill elements are correct, so you can make payment. The public chain and alliance chain node operation agreement maintains the operation of the Internet of Things system. The alliance chain node generates new blocks at a time interval of 30 s. When the node cannot complete the block generation within 30 s, it will rotate to the next node.

## 2. Dynamic Coding Service System

*2.1. Internet of Things.* The IoT paradigm is expected to completely change the way we live and work through a large number of new services based on the seamless interaction between a large number of heterogeneous devices. After decades of creation of the concept of the Internet of Things, in recent years, various communication technologies have gradually emerged, reflecting the diversity of application fields and communication requirements. At present, this heterogeneity and fragmentation of the connectivity landscape hinder the full realization of the vision of the Internet of Things by posing some complex integration challenges. In this case, the emergence of 5G cellular systems with truly ubiquitous, reliable, scalable, and cost-effective connection technologies is considered to be a potential key driver of the yet-to-be-emerging global Internet of Things. Similar to how humans use the Internet, devices will become the main users of the Internet of Things (IoT) ecosystem. Therefore, device-to-device (D2D) communication is expected to become an inherent part of the Internet of Things. Devices will automatically communicate with each other without any centralized control and cooperate in a multihop manner to collect, share, and forward information. The ability to collect relevant information in real time is the key to leveraging the value of the Internet of Things because such information will be transformed into intelligence, which will help create a

smart environment. Ultimately, the quality of the information collected depends on the intelligence of the device. In addition, these communication devices will operate with different networking standards and may encounter intermittent connections with each other, and many of them will be limited by resources. These features bring some networking challenges that traditional routing protocols cannot solve. Therefore, devices will need intelligent routing protocols to be intelligent. Nowadays, the development of traditional business models has become more and more mature, and people use them to guide various e-commerce activities [5, 6]. The Internet of Things (IoT) is an innovative revolution on the Internet and has become a new platform for e-commerce [7]. The flow of static timing constraints is shown in Figure 1.

**2.2. Positive Impact of Digital Currency.** For many people, the concept of digital currency is abstract and confusing. Having confidence in intangible assets without government or precious metal support is a daunting task. However, the rapid spread of smartphones and tablets, the rapid transformation of cross-border banking, and the emergence of non-card real-time payments have made digital payments commonplace. First of all, it has a positive impact. Digital currency creates a relatively novel concept and model, which can improve transaction convenience and reduce transaction costs, promotes the progress and development of shared finance, and decentralizes the digital currency embedded payment system mainly from its own; the system allows users to directly carry out peer-to-peer transactions without resorting to financial units, which can improve transaction efficiency and reduce transaction costs. At the same time, its lower transaction costs will have an impact on traditional payment systems and promote banks and other financial institutions to continuously improve their service quality and reduce transaction costs. Secondly, digital currency can allow people who have not created an account in a financial unit to carry out noncash payments, and the speed is very fast, and the cost is low. Therefore, digital currency has a positive impact on the popularization of finance, especially for those who are relatively backward in finance. In regions and countries, the benefits it brings are also very large [8, 9]. Digital currency can realize network transfer with the help of mobile phones and so on, and the recipient only needs to obtain digital currency to exchange activities with it [10]. When there is no code perception, it is shown in Figure 2, and when there is code perception, it is shown in Figure 3.

**2.3. Verify Algorithm.** The Verify algorithm uses a layered verification mechanism to verify whether the hidden transaction amount is correct [11, 12]. Specifically, it will be stated in Verify-I that the payer has enough bitcoins; that is, the payee must be convinced that the input of the transaction

is greater than the output. Then, the recipient verifies that the transaction amount promised by the wallet is equal to the actual transaction amount in Verify-II. In the commitment phase, the wallet will make a commitment to both the bitcoin deposit amount  $b_1$  and the bitcoin transaction amount  $b_2$  ( $b < 2 \leq b_1$ ). In addition, the wallet will also promise the difference between  $b_1$  and  $b_2$ . In the verification phase, the correctness of the commitment value is checked in two steps: (1) the difference between the input amount  $b_1$  and the output amount  $b_2$  is always positive and (2) the promise made by the payee to the correct transaction amount is always the same as the promise value  $c$  of the wallet [13]. Online wallet calculates  $F_1$  and  $F_2$  and sends  $(F_2, F_3)$  to the payee:

$$F_1 = h_1^{m-a} h_2 \bmod n,$$

$$u(t) = u_i(t - \Delta t) R^L(\Delta t) + u_i(t) \sum_{i=1}^n (t + \Delta t)(t - \Delta t),$$

$$e_0(\lambda, T) = \frac{5\pi HC}{\lambda^6} \frac{1}{e^{(hc/\lambda KT)}} - \frac{hc}{e}. \quad (1)$$

Recipient calculation is as follows:

$$U = \frac{h_1}{F_3} = h_1^\beta h_2^{-r} a^{2-q_1} \alpha^{-q_2} \bmod n,$$

$$F_2 = \frac{1}{Q} \sum_{\omega \in Q} \exp\left\{-\frac{U(z)}{T}\right\} \delta(z - \omega), \quad (2)$$

$$M = T[x, y, p(x, y), f(x, y)].$$

If the check passes, the recipient can trust  $m \geq \alpha$ . At this stage, my country's main basis for this supervision is only this notice, and there is no other effective document [14, 15]. Moreover, the measures of my country's regulatory authorities are mainly aimed at preventing financial risks and money laundering crimes. The protection of financial consumers is mainly based on the principle of "risk yourself." The overall rules are relatively rough. The filing system for trading platforms is not a licensing system. This has led to the obligation of financial institutions and digital cryptocurrency trading platforms to be limited to popularizing industry knowledge and risk disclosure, accepting irregular administrative inspections, and so on [16] and make substantive requirements for deeper content such as the entry barriers of the trading platform, network security, information disclosure, and fund management. As a result, the chaos in the digital cryptocurrency market has not been curbed, and even with the emergence of ICOs, lawsuits for digital cryptocurrency continue to increase [17, 18]. Recipient calculation is as follows:

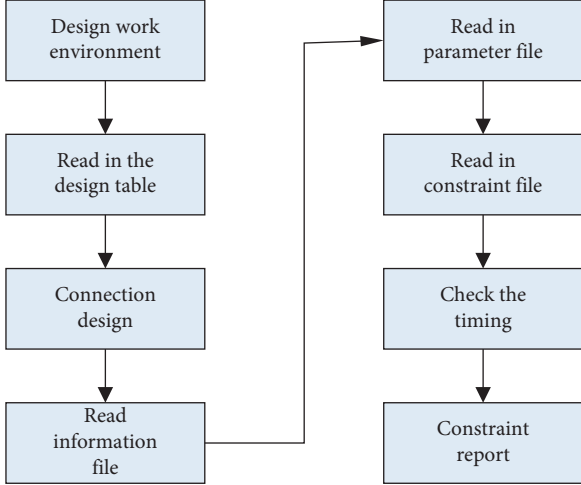


FIGURE 1: Flow of static timing constraints.

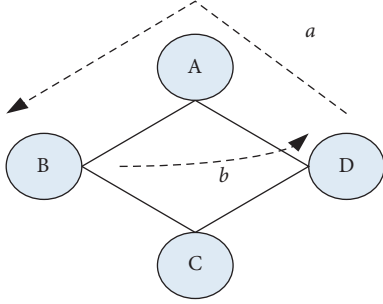


FIGURE 2: No coding perception.

$$\begin{aligned}
 c &= \frac{c_1}{c_2} = h_1^{b_1 - b_2'} h_2^{r_1 - r_2}, \\
 c_1 &= -2 \left( r_{ui} - \sum_{f=1}^F p_{uf} q_{if} \right) + 2p_{uf}, \\
 c_2 &= \frac{\sum_{i \in L} (h_1 - \bar{r}_u) \times (h_2 - r_u)}{\sqrt{\sum_{i \in L} (r_{ui} - \bar{r}_u)^2 \sum_{i \in L} (r_{vi} - \bar{r}_v)^2}}
 \end{aligned} \quad (3)$$

Verify that  $c$  is equal to  $c'$ . If there is any error in the verification process, the payee returns 0 and rejects the transaction [19]. If the verification is successful, the payee returns 1, and then the online wallet broadcasts the encrypted transaction and sends the digital currency amount to the payee account [20, 21].

**2.4. Code-Aware Routing Algorithm.** In wireless networks, the quality of the channel is the key to the successful transmission of data, which is generally measured by the signal-to-noise ratio [22]. The channel quality has great

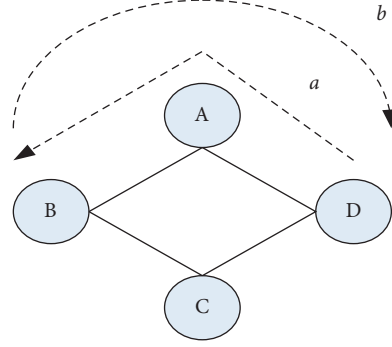


FIGURE 3: Coding-aware.

differences in different links of the actual network [23, 24].

$$\begin{aligned}
 h &= \sum_{i=1}^n c o_i h_{\min} + \sum_{i=1}^n c o_i, \\
 D_1^K &= P(x_i \in x_k, y_i = 1) = \sum_i x \in x_k(i), \\
 D_{-1}^K &= P(x_i \in x_k, y_i = -1) = \sum_i x \in x_k D_{-1}.
 \end{aligned} \quad (4)$$

The final strong classifier is as follows:

$$H(x) = \sin g \left( \sum_{t=1}^T h_t(x) - \text{threshold} \right). \quad (5)$$

If you want to transmit under poor channel quality, you must choose a relatively low transmission rate. There are many broadcasting situations in network coding. When broadcasting to various downstream nodes, the signal-to-noise ratio of each single link in the broadcasting link is different [25]. Choosing a suitable broadcasting transmission rate can make the throughput efficiency reach during this broadcast maximum. The coding-aware routing algorithm based on rate selection is expressed as follows:

$$\begin{aligned}
 M &= \min_{l \in L} \left\{ M_l \mid M_l = \frac{H_l * H_l'}{R_l} \right\}, \\
 \text{MIQ}_s(c) &= \text{MQ}_s(c) + \sum_{t \in c} \text{MQ}_s(i), \\
 \text{MIQ}_d(c) &= \text{MQ}_d(c) + \sum_{t \in c} \text{MQ}(l),
 \end{aligned} \quad (6)$$

$$\text{CRM}_j = \frac{1 + \text{MIQ}_d(l)}{1 - P_t}.$$

Among them,  $M$  is a unicast or broadcast link on the entire link  $L$  for data packet transmission. When the data

link layer successfully transmits a data packet, the routing layer initiates a transmission [26].

$$\text{Angle}(a, r) = \arccos\left(\frac{a \bullet r}{|a| * |r|}\right),$$

$$\min(E^2) = \sum_{k=1}^N [y'(k) - y(k)]^2, \quad (7)$$

$$s = \{x_1, y_1\}, (x_2, y_2), \dots, (x_N, y_N).$$

The probability of successful data packet transmission on the data link layer is as follows:

$$p_i^j = \sum_{k=1}^5 (1 - p_{1i}^j)^{k-1} p_{1i}^j,$$

$$\text{CRM}_L = \sum_{l \in L} \text{CRM}_J,$$

$$F^2(s, v) = \frac{1}{s} \sum_{i=1}^s \{Y[N - (v - N_s) + i] - y_v(i)\}^2, \quad (8)$$

$$P_{n,j} = \frac{P_{n,t}}{P_{n,t-1}} - 1,$$

$$P_j = \frac{P_{n,t}}{P_{n,t-1}}.$$

That is, the probability of starting a routing layer transmission is  $p_i^j$ . The process of the Verify algorithm generating code is shown in Figure 4.

### 3. Dynamic Coding Service System Experiment

**3.1. System Framework Design.** The system is divided into four parts: the background server, the foreground management system, the foreground business system, and the cipher. The back-end server is equivalent to a certification authority CA, which stores the user's certificate information and completes the actual verification process. It is generally the server of the head office. The front-end management system is the management program on the front-end computer of the branch, which manages the user account information, the issuance and management of the user password, the management of the operator and the log query, and so on. The front desk business system is a business management program on the front-end computer of the branch. It uses the information on the check and the payment password to verify the authenticity of the check and generally provides services for the bank transfer system. The cipher is a handheld device used by the user, which is issued to the user by the bank to implement the user-side algorithm in hardware, including the generation of the key and the generation of the payment password. The system framework is shown in Figure 5.

**3.2. Cipher.** The encryption chip integrates RSA and SHA-1 algorithms and saves the user's key. There are two types of chips: A slice is mainly a public key algorithm, which is used

in the payment password verification subsystem, and B slice is mainly a private key algorithm, which is used in the payment cipher used by the account opening unit. The payment cipher adopts the B-chip arithmetic chip to make a handheld device to manage the user's account and key and generate payment password. The payment cipher is managed by the bank. When the user opens an account, the bank registers and sends it to the user, and the key is generated by the algorithm chip, and the public key is stored in the bank's back-end server.

**3.3. Password Management.** The payment cipher is issued by the bank to the customer who opens an account with the bank and downloads the account number of the customer with the bank and the corresponding account key in it. When the customer uses the account to issue a bill, enter the bill number, bill type, amount, and other information on the payment cipher and use the payment cipher to automatically calculate a string of numbers. This number is closely related to the payer account number, receiver account number, bill type, bill number, amount, and date of signing and is called the payment password. The customer fills in the number on the bill as the digital seal of the bill. When the bill is delivered to the bank, the bank enters the various elements of the bill on the counter terminal and transmits it to the verification machine through the bank network for verification. If the verification is correct, it indicates that the bill is indeed issued by the customer, and the bill elements are correct, so you can make payment.

**3.4. Calculate the Plaintext Format of Payment Password.** The format of the 48-byte plaintext data PLAIN\_TXT input to the B slice when the payment password is generated is shown in Table 1. After the chip output is converted, it is a 19-digit integer plus 1 identification bit to form a 20-digit payment password output. The plaintext format of the payment password is shown in Table 1.

**3.5. Design of the Alliance Chain.** System initialization completes the generation of system parameters and the initial state of the blockchain; transaction verification and forwarding are the process of sharing information among alliance chain members, ensuring that nodes reach consensus on the same basis; the consensus process includes the specific process of node interaction in CPBFT; transaction confusion is responsible for after the transaction is confirmed, and the transaction is processed before being sent to the public chain node to remove the transaction relationship information to protect transaction privacy; the final transaction traceability is the process of internal supervision of the system. The communication between nodes in the alliance chain uses encrypted channels to prevent information leakage when the communication transmits the plaintext and completes transactions.

**3.6. Choice of Algorithm.** The Verify algorithm aims to check that the difference  $b1 - b2$  between the input amount and the

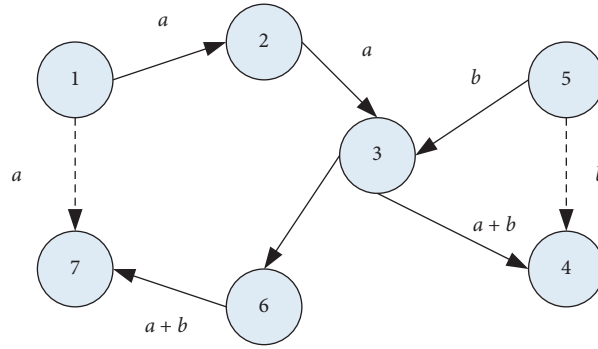


FIGURE 4: The process of the Verify algorithm to generate code.

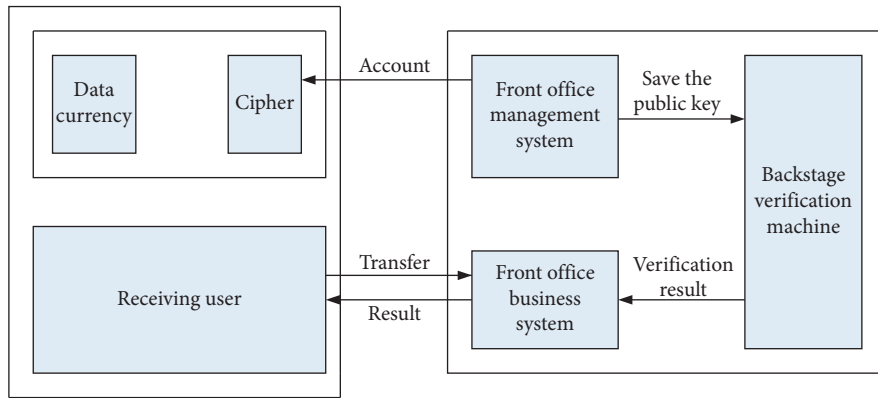


FIGURE 5: System framework.

TABLE 1: Plaintext format of the payment password.

|            |    |  |
|------------|----|--|
| ACCU       | 16 | Account, compressed BCD code               |
| Service    | 4  | Business type, $0 \times 31 - 0 \times 35$ |
| Date       | 4  | Date, compressed BCD code                  |
| Ticket_num | 8  | Voucher number, compressed BCD code        |
| Balance    | 9  | Amount, in minutes, compressed BCD code    |

output amount of the transaction is always positive, and the output amount  $b_2$  is the transaction amount specified by the payee. The algorithm first confirms that the promised secret value is always positive. To this end, the payer commits to the difference between the input amount and the output amount of the transaction. Then, use the range proof method to convince the payee to believe that the promised secret value is always positive. Secondly, because the recipient knows the correct transaction amount  $b_2$  in advance, he also needs to use the correct transaction amount and some auxiliary evidence  $c_1$  and  $r_2$  to verify that the final commitment value is equal to the commitment value  $c$  made by the wallet. We call this two-step verification method a layered verification mechanism. If the algorithm returns 1, the protocol goes to the next step.

**3.7. Operation Process Design.** The peer relationship between public chain nodes and alliance chain nodes in the system is just different in processing messages. Together, these two

parts can be regarded as servers. The user's wallet is the client and sends operation requests to the server. After each part of the system starts running, it performs related functions according to the protocol designed in chapter 4. When there is no user to send a transaction, the blockchain node automatically runs the consensus process in a loop to maintain the consistency of the system and waits to process the sent transaction information. After the user sends the transaction, it is collected and processed by the alliance chain nodes, including verification transactions, packaged transactions, and transaction confusion, and finally the confirmed transactions are confused and broadcasted to the public chain nodes to complete the complete transaction confirmation and recording process. When the system initiates the traceability, the entire process only occurs between the alliance chain nodes. The traceability initiating node initiates a request to other alliance chain nodes including the supervision node. The other alliance chain nodes in the figure include more than one node, and the request is judged separately, and the supervised node obtains the user identity after recovering the

key. The design values of running nodes are shown in Table 2. The running process is shown in Figure 6.

**3.8. Implementation of Routing Algorithm.** The route discovery process of RDCAR is as follows:

- (1) The source node initiates the route discovery process of the wireless network by broadcasting a route request (RREQ) message. In our algorithm, the routing request message needs to include the following messages: neighbor nodes within the range of the source node, high channel quality signal-to-noise ratio, and the path that has been transmitted.
- (2) The intermediate node receives the RREQ message; first to avoid loops, check whether the same RREQ message has been introduced and discard it if it is received; otherwise, it will buffer the message and attach its own neighbor node list, corresponding to the signal noise of the channel link, and then update the RREQ message and broadcast it.
- (3) When the RREQ arrives at the destination node, the destination node sends a route reply request (RREP) message to the source node. This is a unicast message that contains information on this unicast path.
- (4) When the intermediate node receives the RREP message, it selects a value from the set of rates supported in the 802.11 protocol, calculates the expected number of transmissions based on the stored signal-to-noise ratio information, selects the appropriate rate, updates the neighbor node list, and finally calculates the smallest metric value. Then, add this information to RREP. Take out the RREQ path in the cache and compare it with the upstream path set in the RREP. Use the conditions of the encoding node above to check whether there is an encoding opportunity for the new stream. If it exists, calculate the metric value of the broadcast link. Add the minimum metric value to RREP, and continue forwarding the cached path until it reaches each source node.
- (5) Maintain routing regularly, update routing messages, and reselect the most appropriate path for the current distributed flow. The realization of the routing algorithm is shown in Figure 7.

**3.9. IoT Node Operation.** The public chain and alliance chain node operation agreement maintains the operation of the Internet of Things system. The alliance chain node generates new blocks at a time interval of 30 s. When the node cannot complete the block generation within 30 s, it will rotate to the next node. Similar rules are also used in the public chain to generate block rights maintenance system operation.

**3.10. Realization of User Transfer Process.** The transfer behavior takes place in the user's wallet, and the user first verifies his identity using the wallet. The user's login name and corresponding password are stored locally, and the

password retains the value after MD5 calculation. During the login process, compare whether it is the same as the previously saved record. After the user logs in, the wallet will check whether the user has an identity certificate issued by the authentication server in the system. Only transactions sent by users registered in the system will be received by the alliance chain node. When the user does not store the identity certificate in the node, the wallet will prompt user complete the authentication. Users can use the wallet to initiate transfers, and they can also query the status of the account. Transfers between users need to use the public key as the address to receive the transfer. In fact, the public key is associated with the payment transaction so that the currency obtained in the payment transaction is used. When making a payment, you can prove your ownership of the currency through the private key corresponding to the public key and pass the cryptographic algorithm without the need to pass a centralized credit institution. The generated payment address will be displayed, and the corresponding private key will be stored in the file by the wallet and used in the transaction. The payment address can be sent to other users in the digital currency system by means of communication outside the system.

After sending the transaction, the user needs to wait for the alliance chain node and the public chain node to perform a series of operations and finally store the relevant transfer-in and transfer-out transactions in the public chain block. The wallet has stored the label and content of the complete transaction for a period of time. By comparing the data stored in the public chain, you can check whether the transaction you initiated has been written to the block and confirmed by multiple blocks. To initiate a transfer, users need to use their USOT as the payment method, provide the corresponding private key, send currency to the public key provided by the other party, and provide one or more public key addresses of their own as the change address. The total amount of currency in the USOT provided by the transaction initiator cannot be less than the payment amount. If the balance cannot be exactly equal, the excess amount will be sent to the change address and returned to the payer.

**3.11. Traceability and Transaction Disclosure.** Traceability is initiated and completed internally by the alliance chain, and the result of the transaction traceability can be seen in the end. The goal of retroactive transactions is generally a transfer-out transaction because the initiator of a complete transaction on the transfer-out exchange is the owner of the transfer-out transaction address, and the identity of the initiator can be queried. If no one spends it for a transfer-in transaction, it cannot get the owner of the transaction. When searching for a transaction, according to the query transaction serial number, it traverses the plaintext of transaction information saved in the block, finds the block containing the same serial number, decrypts the transaction, finally finds the transaction, and displays the complete record in the transaction.

**3.12. System Test and Implementation.** This system runs under Windows and uses Java language for programming.

TABLE 2: Design values of operating nodes.

| Constraint content    | Settings (ns) |
|-----------------------|---------------|
| Input delay           | 3.1           |
| Output delay          | 3.0           |
| Clock jitter and skew | 3.2           |
| Drive capability      | 3.2           |

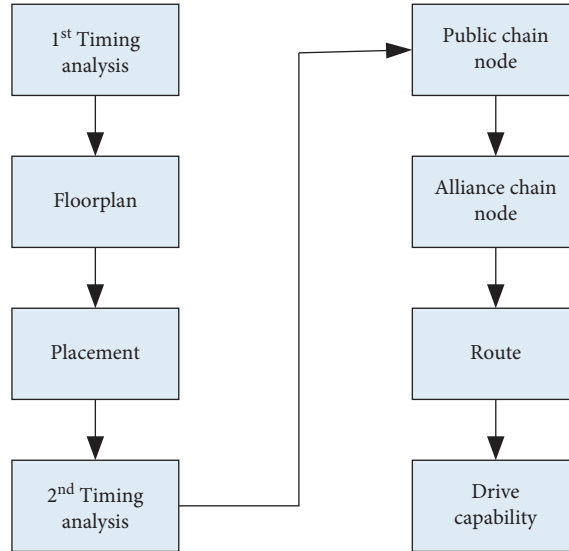


FIGURE 6: Running process.

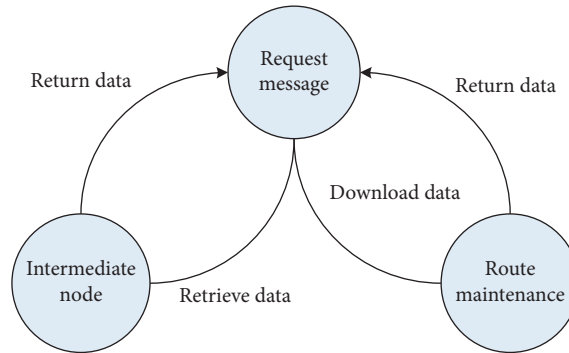


FIGURE 7: Implementation of the routing algorithm.

The machine configuration for system testing is shown in Table 3. This system uses the JPBC (Java Pairing-Based Cryptography) library. JPBC is a Java package based on the paired cryptography library function (PBC). JPBC completely breaks away from the dependence of PBC, realizes the pair operation completely based on Java, and puts aside the limitations of the platform itself. This system uses it as the basic support library for cryptography. In addition to the basic upload and download function files, this system mainly includes the following types of files that generate user keys, generate keyword ciphertexts and secret doors, and implement keyword encryption and search. CLPEKS.java is used to implement various algorithms. CLPEKS Pub Params.java is used to store public parameters, Service and Client Key Interface java is an interface class for client and

server keys, CLPEKS Secret Key.java and CLPEKS Client Key java store server and client keys, respectively, and CLPEKS Cipher text.java and CLPEKS Trap door.java store ciphertext and secret door, respectively.

#### 4. Analysis of Dynamic Coding Service System

**4.1. Algorithm Performance Analysis.** The performance test mainly includes two aspects. On the one hand, the algorithm performance changes under different matrix sizes, that is, when  $N$  is different; on the other hand, the performance differs between the test algorithm and the SM2 algorithm. First, test the performance of the algorithm when different matrix sizes are different, that is, when  $N$  is different. When the test  $N$  is 4, 8, 16, 32, and 64, the number of signatures and

TABLE 3: Machine configuration during the system test.

| Project          | Value  |
|------------------|--|
| Operating system | Microsoft Windows 7 Ultimate                 |
| Version          | 6.1.7601 Service Pack1 internal version 7601 |
| System type      | x64-based PC                                 |
| Processor        | Intel8CoreTM i3-2310M                        |
| CPU              | 2.10 GHz                                     |
| RAM capacity     | 8 GB   |

verifications per second by the algorithm and the experimental results retain two significant digits. The performance of the algorithm at different matrix sizes is shown in Table 4. It can be seen from Table 4 that with the increase in the matrix size, the performance of algorithm signature and verification is gradually decreasing. The reason is that the larger the matrix size, the more cycles in the GeneratePrivateKey and Public Key Derivation GeneratePublicKey methods. At the same time, the performance of the verification algorithm drops faster because the loop of the GeneratePublicKey method in the verification process contains the accumulation of elliptic curve points which takes longer. However, although the performance of the algorithm decreases as the matrix size  $N$  increases, the security of the algorithm will increase as  $N$  increases. Next, test the performance difference between this algorithm and the standard SM2 signature algorithm. When the test matrix size  $N=64$ , the number of signatures and verifications per second of this algorithm and the SM2 algorithm are the same. The initialization of the private key matrix SKM and the public key matrix PKM will only be performed once, so this part of the time is not calculated. From an algorithm perspective, the performance difference between the two algorithm signatures (or signature verification) is the time it takes to derive the key from the matrix. The performance difference of the signature of the two algorithms is greater than the performance of the verification. This is because the calculation of the private key corresponding to the public key in the private key derivation algorithm (the private key class contains the public key) will involve the elliptic curve dot product operation. The dot multiplication operation takes a long time. It can be seen from Table 4 that when the matrix size is  $N=64$ , the signature and verification efficiency of this algorithm are both close to half of the efficiency of the SM2 algorithm, and the performance difference is due to the secret key of the proxy signature private key and the proxy verification public key derived from overhead. On the whole, performance has been lost, but according to the safety analysis of this algorithm, the overall safety limit of the system has been increased from 1 to 33. After that, the SM2 hash algorithm and this research algorithm are tested for performance comparison. Use the mkfile command to generate 16B, 1KB, 1MB, and 1GB files in sequence as input. After 10 tests, according to the test results, as the length of the input message increases, the computational efficiency of SM2 and this research show an upward trend first and then reach a peak. The peak algorithm speed of SM2 is around 100 MB/s, and the peak algorithm speed of RDCAR is around 370 MB/s. On the whole, the efficiency of the algorithm in this study is higher than that of SM2. This is

not because of the gap in the algorithm itself but because the bottom layer of the Go algorithm is implemented in assembly, which is more efficient. The performance analysis of the algorithm at different matrix sizes is shown in Figure 8. Before the simulation is carried out, it is necessary to specify specific voltage values for the power signal and ground as shown in Table 5.

*4.2. Dynamic Coding Analysis.* Use simulation tools to simulate the RDCAR routing algorithm. We take the network model of the grid graph and distribute 25 nodes in the range of  $100 \times 100 \text{ m}^2$ , and the communication range of each node is 20 m. 11 streams are randomly generated in the network. The purpose of our experiment is to compare RDCAR, DCAR protocol, and COPE coding protocol. Since there is no DCAR protocol and COPE coding protocol, there is no rate-aware algorithm, so the rate of DCAR and COPE cannot be automatically adjusted. Let's take a look at the coding opportunities and throughput. Under different speeds and signal-to-noise ratios, compare the coding opportunities and throughput of RDCAR. The rates of COPE and DCAR we used here are the same, so the difference in coding opportunities is enough to reflect the difference in throughput. At the same time, the increase in coding opportunities will lead to the increase in throughput, so the number of coding opportunities can reflect the size of throughput. The simulation shows the comparison of coding opportunities at different rates in a low channel capacity environment. We can see that in the actual situation of about 5 to 10 dB, the channel quality is very poor and the packet loss rate is quite high. DCAR and COPE use a minimum of 12 mbps, and DCAR and COPE use 24 mbps for coding opportunities. Our RDCAR uses a rate selection algorithm, so it will use a lower transmission rate to ensure the number of successes on a poor link and use a higher transmission rate on a better link to increase throughput. We found that DCAR adopts coding-aware routing scheme and can actively discover coding opportunities. Therefore, under the same circumstances, no matter what transmission rate is used, it is better than COPE. When the channel environment of 5 to 10 dB is poor, compared with the two cases of 12 mbps and 24 mbps, using 12 mbps to guarantee the transmission rate, the coding opportunity and throughput are better than those of 24 mbps. The medium quality channel ranges from 10 to 20 dB. Since under medium channel quality, a certain transmission success rate can be guaranteed even when 48 mbps is used again, and we can see that as the transmission rate increases, under the premise of ensuring a



TABLE 4: Algorithm performance at different matrix sizes.

| Algorithm | $N=4$  | $N=8$  | $N=16$ | $N=32$ | $N=64$ |
|-----------|--------|--------|--------|--------|--------|
| Signature | 221.72 | 262.47 | 251.89 | 239.81 | 219.78 |
| Check     | 245.10 | 227.27 | 224.72 | 189.39 | 155.28 |

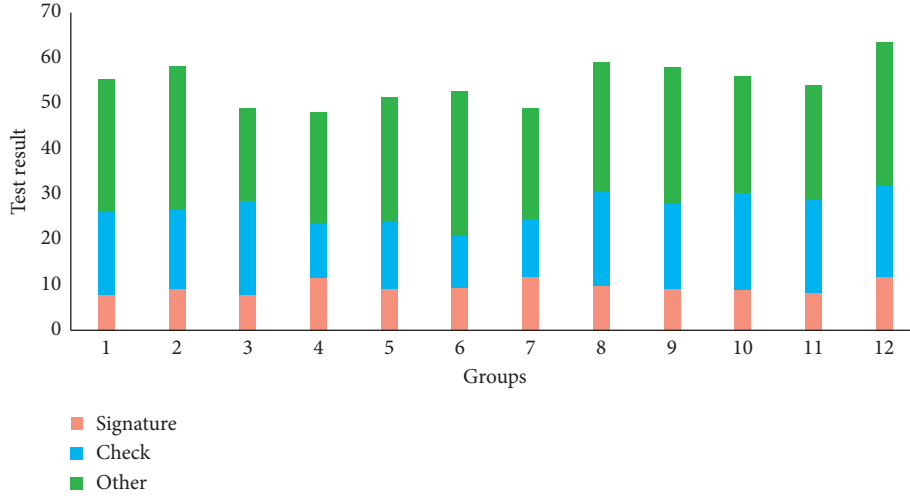


FIGURE 8: Algorithm performance analysis at different matrix sizes.

TABLE 5: Before the simulation, you need to specify specific voltage values for the power signal and ground.

| Net name | Voltage value (V) |
|----------|-------------------|
| AWCC1V8  | 1.8               |
| DD RYTT  | 075               |
| GND      | 1.2               |
| YCC1V2   | 1.8               |
| YCC1V8   | 33                |

certain transmission probability, the coding of COPE and DCAR opportunities and throughput gradually approach RDCAR. Due to the high channel quality, the success rate of information transmission is very high, and the main factor that affects coding opportunities and throughput is the transmission rate. Therefore, we can see that in Figure 9, when DCAR and COPE only use low-speed channels, the low transmission rate guarantees success rate is no longer applicable but greatly affects the transmission efficiency, making the coding opportunity and throughput smaller. DCAR and COPE use medium-speed channels. It can be seen that the coding efficiency of RDCAR is still far behind when using 24 mbps. The transmission rate of RDCAR in our actual simulation is about 48 mbps. Therefore, we found in subsequent experiments that if 48 mbps high-speed transmission is used, the coding opportunities of DCAR and RDCAR are quite close, and there is a certain gap between COPE and DCAR. The coding situation monitored by different transmission rates is shown in Table 6. The learning step length of the RDCAR algorithm is shown in Table 7. The original high-frequency coded signal collected by using the oscilloscope is shown in Figure 10.

*4.3. Function Analysis.* The model proposed in this study not only includes the antitampering, traceability, and decentralization characteristics of transactions in the existing digital currency system but also adds supervisable attributes to the system to enhance the system's ability to protect user privacy. The performance test is divided into two aspects. On the one hand, it tests the performance changes of the algorithm under different matrix sizes, that is,  $N$ ; on the other hand, it tests the performance comparison between this algorithm and the standard SM2 signature algorithm. During the test, first create a plaintext byte array and assign values, then create a key byte array and assign values, create an encryption instance by calling `sm4.NewCipher(key)`, and then call `c.Encrypt` to calculate the encryption result. Then, call `c.Decrypt` to calculate the decryption result. The performance of the algorithm changes when the test matrix size  $N$  is different. When the test  $N$  is 8, 16, 32, and 64, calculate the number of signature tests and verification times of the algorithm per second. The result of functional analysis is shown in Figure 11. Since authorization generation and authorization verification will only occur once when the system is initialized, this part of the time is not calculated. In

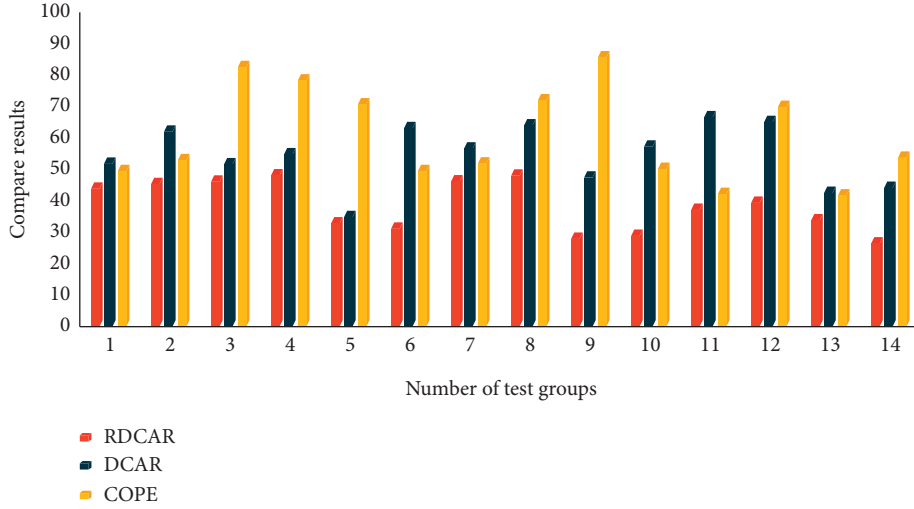


FIGURE 9: Comparison of encoding opportunities at different rates.

TABLE 6: Coding situation monitored by different transmission rates.

| Project | Infeasible area    | Feasible region | Near real POF      |
|---------|--------------------|-----------------|--------------------|
| 5 dB    | (0.10, 0.20)       | (0.40, 0.55)    | (0.30, 0.45)       |
| 10 dB   | (0.10, 0.10)       | (0.70, 0.70)    | (0.60, 0.64)       |
| 12 mbps | (0.10, 0.20)       | (0.40, 0.50)    | (0.24, 0.27)       |
| 24 mbps | (0.10, 0.20, 0.10) | (30, 0.7, 0.25) | (0.12, 0.31, 0.67) |
| 36 mbps | (0.20, 0.50, 0.60) | (70, 0.7, 0.50) | (0.50, 0.77, 0.37) |
| 48 mbps | (0.20, 0.50, 0.60) | (70, 0.7, 0.50) | (0.50, 0.77, 0.37) |

TABLE 7: Learning step size of the RDCAR algorithm.

| Number of learning samples | Number of test samples | Number of hidden nodes | Learning step |
|----------------------------|------------------------|------------------------|---------------|
| 100                        | 50                     | 8                      | 0.005, 0.005  |
| 100                        | 50                     | 8                      | 0.005, 0.005  |

terms of privacy protection, the complete transaction records are encrypted and stored. Only after initiating the traceability of the transaction and voting by the members of the alliance chain, can the transaction plaintext be viewed by the special members of the alliance chain. After completing a consensus, the alliance chain broadcasts a batch of transfer-in and transfer-out transactions, hiding multiple pieces of information belonging to the same complete transaction in a large amount of similar information and avoiding the leakage of user privacy in the process of information interaction. The public transaction records stored in the public chain for query, because of truncation and confusion, lose the traceable transaction relationship information and have the unlinkability of the transaction output required by the digital currency public chain data. The data stored in the alliance chain node have the most restrictive restrictions on access rights. During the normal operation, the alliance chain node only has the write permission, and the verification function in the system is completed by the data in the

public chain block. In terms of supervisable attributes, when there is a need for the system to trace the transaction, the alliance chain node initiates a transaction traceability application, and the nodes can understand the reason for initiating the traceability outside the system, and then the supervisory authority node is responsible for restoring the key and decrypting the traceability. Transaction records finally obtain the identities of relevant transaction participants through the identity verification server. But, only after a USOT is spent, the consortium chain can obtain its owner's identity signature through a transaction request from a public chain user. If a USOT has not been spent, its owner cannot be known. In order to increase the supervisable properties of the system, this research adds three parts of the supervision agency, identity authentication server, and encrypted storage to the Internet of Things system. When there is no need to initiate traceability, the supervisory agency only participates in the consensus process as a participant in the alliance chain, and only when it needs to

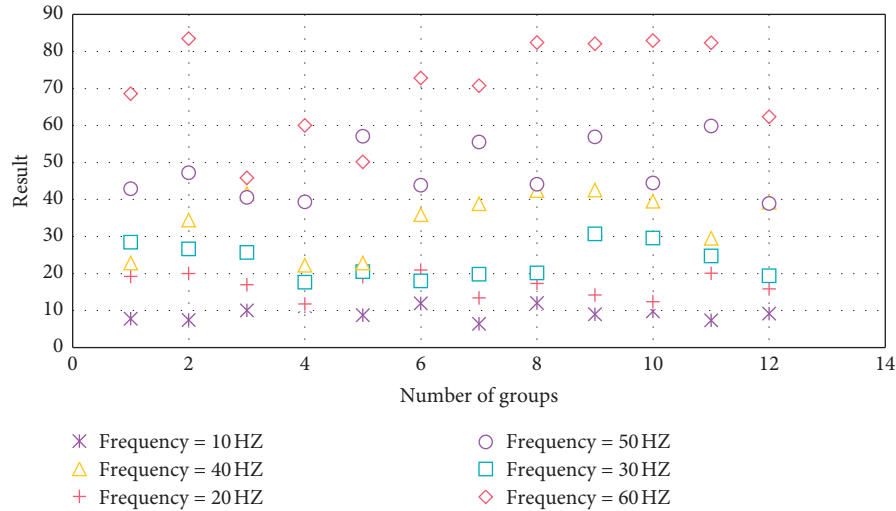


FIGURE 10: The original high-frequency coded signal collected by using an oscilloscope.

trace the transaction, can it act as a trusted secret shared share collector to decrypt data. The identity authentication server issues identity certificates to registered users, which can trace the transaction to a clear user and achieve thorough supervision. The addition of supervisory attributes reduces the degree of decentralization of the system, but it does not damage the Internet of Things from the user's perspective. The original intention of the design is that the members and data of the alliance chain and the public chain do not rely on centralized credit, and it can also achieve better protection for users. When a user loses an account or transaction certification document such as an identity certificate, he can protect his account through the identity authentication server and the system. Encrypted storage is to prevent consortium chain members from leaking transaction information after being attacked. The function of the regulatory agency has affected the decentralized structure, and it has also become a security weakness that the system may be attacked. The use of secret sharing to a certain extent prevents the possibility of stealing transaction data with the regulatory agency as the target, and only alliance chain members agree to restore only when the key is the supervisory authority which has the ability to decrypt data. The test results after changing the maximum number of learning times are shown in Table 8. The regulatory error is shown in Table 9.

**4.4. Security Analysis.** The issuance of legal digital currency requires the transformation of the payment system infrastructure. As a digital form of currency, legal digital currency needs to adopt extremely high technology in any link of circulation to reduce the degree of operational risk. Once the legal digital currency infrastructure is destroyed, it will cause the entire financial system to suffer losses. The system security results are shown in Figure 12. The analysis of legal digital currency has to deal with a large number of transactions, and the most mature distributed ledger technology cannot fully meet the requirements of the central bank's

payment system. In its fiat digital currency project, the Bank of Canada uses distributed ledger technology (DLT) to build a payment system, but the performance of distributed ledger technology is not optimistic. According to the report, it is difficult to use distributed ledger technology for transaction systems to process a large amount of instantaneous transaction data. After the official issuance of legal digital currency, it faces far more requirements than Bitcoin in terms of importance and transaction scale. The payment infrastructure established by the central bank must meet the transaction needs of the society. The issuance of legal digital currency will have a profound impact on the payment system of the entire country. Therefore, the construction of the payment system must consider the requirements of scalability, compatibility, and transaction throughput. This article divides the attacks faced by the supervisable digital currency model into three types according to their sources: attacks from outside the system, attacks from alliance chain nodes, and attacks from public chain nodes. In terms of the most basic security of digital currency transactions, the system uses identity certificates and signatures to ensure that attackers cannot forge identities to steal other people's assets; transactions in the system are based on USOT and use a unique public key as an address. The corresponding private key can be unlocked for payment. The purpose of the attack from outside the system is to destroy the function of the system and make the digital currency system unable to operate normally. The target of the attack may be a node in the system or a communication network. Due to the distributed nature of the blockchain, attacking a node in the system will not affect the operation of the system, but when an attacker has the ability to attack multiple nodes, the alliance chain as the core of the system must ensure that the node that stops working cannot exceed 1/3 of the system. Since the public chain uses DPoS as a consensus mechanism, during the election cycle, those nodes that are made public due to the operation of the system are likely to become targets for system attackers. Therefore, in order to ensure that the system does not stop running, it is necessary to

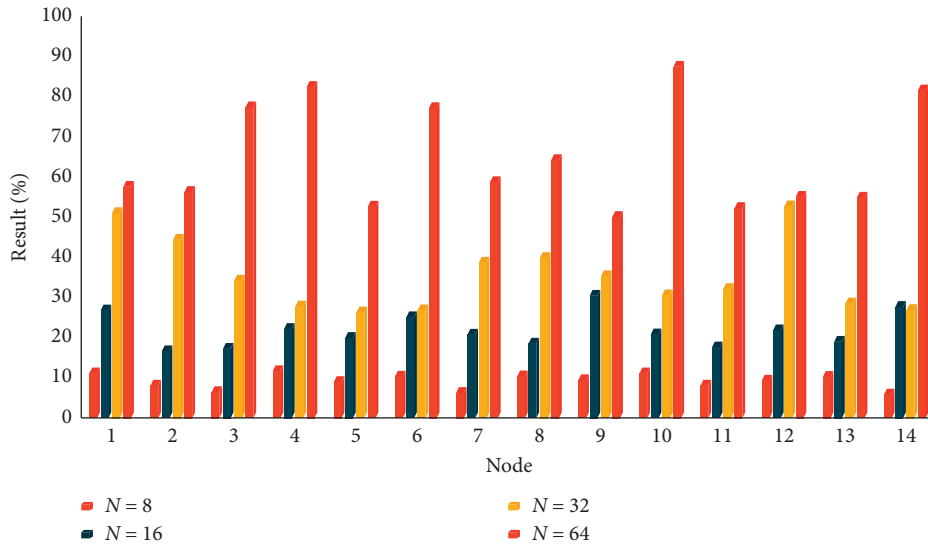


FIGURE 11: Functional analysis results.

TABLE 8: Test results after changing the maximum number of learning.

| Maximum number of studies | Best accuracy rate (%) | Worst accuracy rate (%) | Average accuracy (%) |
|---------------------------|------------------------|-------------------------|----------------------|
| 10000                     | 66                     | 56                      | 60                   |
| 20000                     | 66                     | 60                      | 62                   |

TABLE 9: Regulatory errors.

| Target location                          | Right in front | Upper left | Top right | Right rear |
|--|----------------|------------|-----------|------------|
| Maximum positioning error of X-axis (cm) | 1              | 1          | 1         | 1          |
| Maximum positioning error of Y-axis (cm) | 2              | 1          | 1         | 2          |

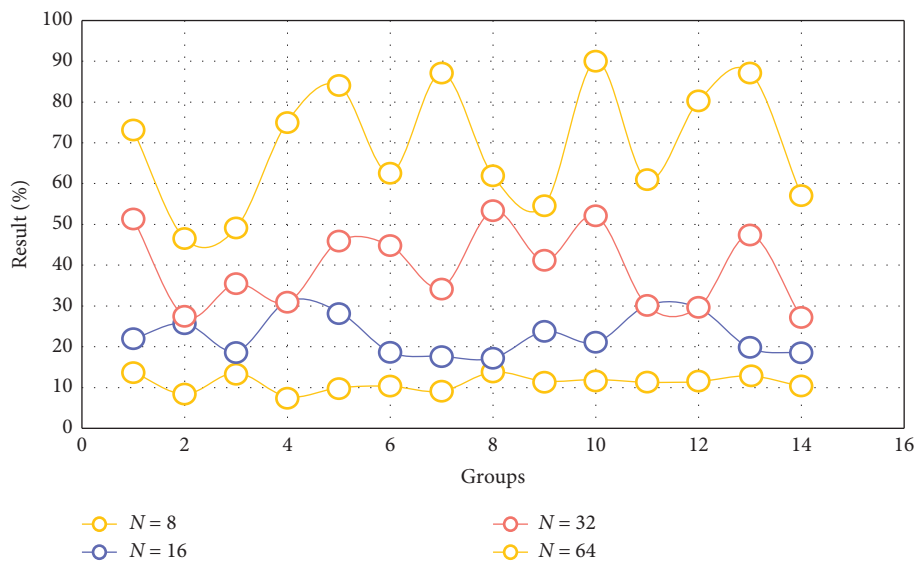


FIGURE 12: System security results.

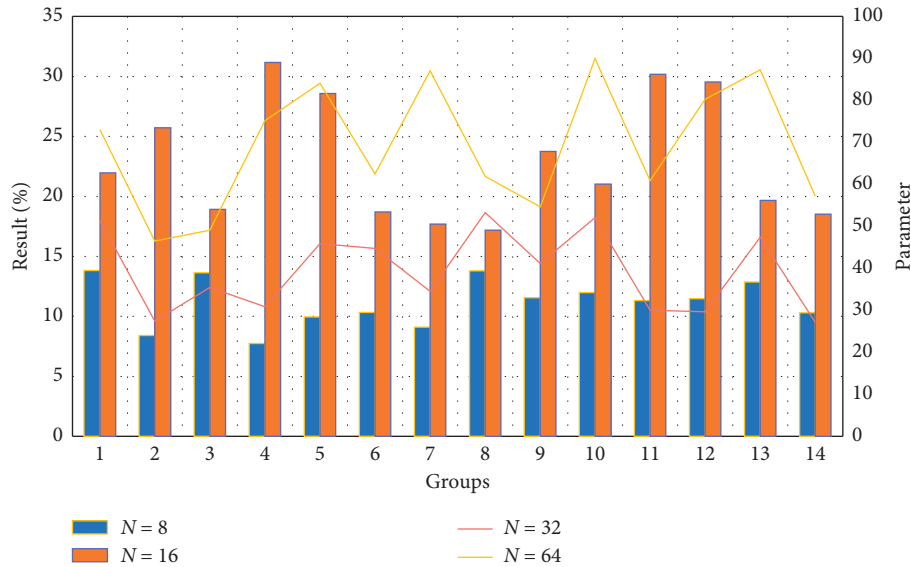


FIGURE 13: Efficiency analysis results.

ensure that the system attackers are all before destroying the current witness node, complete a new round of witness node voting. Attacks from within the system are generally through the creation of system inconsistencies such as the fork of the Internet of Things to achieve double-spending by tampering or canceling transaction records. The CPBFT consensus mechanism adopted by the alliance chain does not have the possibility of forks of the Internet of Things. When the number of colluding attackers does not exceed the threshold, it can prevent double-spending attacks. Although there may be forks in the public chain, only attackers account for more than 50% of the total number of nodes to ensure the success of the attack. The real-name registration mechanism in the background also reduces the possibility of launching and succeeding from inside the system. Digital currency can record and check transaction information and information of both parties, can accurately reflect the implementation of monetary policy, and can strengthen financial management. The transaction record of digital currency cannot be tampered with, can completely record the transactions of each participant, and can form a unified distributed ledger in the entire digital currency system. Through the review of transaction information and the supervision of digital currency circulation, the national regulatory agency can accurately grasp the monetary policy and credit policy in real time, and then scientifically and comprehensively calculate the policy implementation status, and adjust relevant policies in time according to changes in the situation; It can promote the publicity, openness, and integrity of digital currency transactions as a whole by establishing an overlying public credit system.

*4.5. Efficiency Analysis.* Intermediaries such as digital currency trading platforms or traditional financial institutions and certain nonfinancial industries which are likely to participate in the flow of digital currency transactions should keep records and report suspicious transactions.

The specific methods include conducting due diligence on customer identity and storing customer identity information. The public key address, account, transaction nature, date, and amount involved in the transaction are helpful for monitoring transactions, recording suspicious transaction information, and combining the information on the blockchain ledger for more accurate recording. In order to prevent unqualified financial institutions from participating in the payment and settlement system, the access system sets specific conditions so that only payment and settlement participants who meet the corresponding conditions are allowed to be the counterparty of payment and settlement. The efficiency analysis result is shown in Figure 13. As the transparency and intensity of supervision have been greatly improved and the legal digital currency relies on advanced Internet technology, it can better identify the relevant conditions of financial institutions involved in payment and settlement and strengthen prudential supervision. The consensus mechanism used in the public chain part of the system is DPoS, and the consortium chain part uses CPBFT. Both consensus mechanisms are based on voting. The system state is determined according to the choice of the majority of nodes in the system, without the need for additional proof of work. The computational overhead of the system mainly occurs in the verification and encryption of transactions. In the process of verifying the block, each alliance chain node needs to encrypt and compare the transactions packaged into the block. This process uses a public key cryptographic algorithm, which has a high time complexity. Each node has basically the same demand for computing power, and there will be no system security problems and decentralized performance caused by the concentration of computing power. To achieve consistency within the alliance chain, nodes need to broadcast multiple times to achieve information interaction between the two. The communication complexity is  $O(n)$ , where  $n$  is the number of nodes. The system is

TABLE 10: The correlation coefficient matrix of the independent variables.

| X variable<br>Y variable | INF    | LNC    | LNY    | R      | P      | V      |
|--------------------------|--------|--------|--------|--------|--------|--------|
| INF                      | 1.000  | -0.129 | -0.132 | 0.730  | 0.180  | -0.470 |
| LNC                      | -0.129 | 1.000  | 0.999  | -0.473 | -0.363 | 0.876  |
| LNY                      | -0.132 | 0.999  | 1.000  | -0.473 | -0.352 | 0.878. |
| R                        | 0.730  | -0.473 | -0.473 | 1.000  | 0.707  | -0.670 |
| P                        | 0.180  | -0.363 | -0.352 | 0.707  | 1.000  | -0.287 |
| V                        | 0.470  | 0.876  | 0.878  | -0.670 | -0.287 | 1.000  |

TABLE 11: Time series unit root test results.

| Variable            | (c, t, n) | ADF inspection<br>value | Critical<br>value | Conclusion    |
|---------------------|-----------|-------------------------|-------------------|---------------|
| $\ln y$             | (0, 0, 0) | 6.962                   | -1.966            | Smooth        |
| $\Delta \ln y$      | (c, 0, 0) | -2.998                  | -1.956            | Nonstationary |
| Inf                 | (c, 0, 0) | -1.627                  | -1.956            | Nonstationary |
| $\Delta \text{inf}$ | (c, 0, 0) | -3.809                  | -1.956            | Smooth        |
| $p$                 | (c, 0, 0) | -0.182                  | -1.956            | Nonstationary |
| $\Delta p$          | (c, 0, 0) | -3.898                  | -1.956            | Smooth        |

designed so that each participant can control multiple nodes proportionally. The more the number of nodes, the stronger the antiattack ability of the system, but the communication overhead will also increase. The system maintains two blockchains at the same time. Compared with a single chain system, the communication between some nodes of different systems is increased. If the nodes of the alliance chain and the public chain communicate with each other, the complexity of the communication is  $O(n, m)$ , where  $n$  is the number of nodes in the alliance chain,  $m$  is the number of nodes in the public chain, and  $m$  may be multiples of  $n$ , so the pair can be appropriately relaxed. The communication requirements between two blockchain nodes can reduce the communication complexity to  $O(m)$ . The supervisable digital currency system uses CPBFT and DPoS improved in this research. The characteristics of these two consensus mechanisms are that the block generation interval is short and the system transaction throughput is high, so there is no system bottleneck caused by increasing the system scale in the consensus mechanism. The correlation coefficient matrix of the independent variables is shown in Table 10. The unit root test results of the time series are shown in Table 11.

## 5. Conclusion

This research mainly discusses the intelligent digital currency and dynamic coding service system based on the Internet of Things technology. To a certain extent, digital currency can help save currency issuance and circulation costs, improve the effectiveness of monetary policy, accelerate the pace of development to a cashless society, and adopt effective methods for promotion and application. It is expected that digital currency will be widely used by all people, thereby contributing to building a robust and efficient new financial system. At the same time, the current

status of the use of digital currencies at home and abroad is also analyzed. The digital currency market is frequently traded. According to the traditional trading market, digital currency is an active market, but the currency value stability of digital currency is poor.

This research is mainly based on the RDCAR algorithm to realize the traceability of the route discovery process of the wireless network initiated and completed within the alliance chain, and finally the results of the transaction traceability can be seen. The goal of retroactive transactions is generally a transfer-out transaction because the initiator of a complete transaction on the transfer-out exchange is the owner of the transfer-out transaction address, and the identity of the initiator can be queried. If no one spends it for a transfer-in transaction, it cannot get the owner of the transaction. When searching for a transaction, according to the query transaction serial number, it traverses the plaintext of transaction information saved in the block, finds the block containing the same serial number, decrypts the transaction, finally finds the transaction, and displays the complete record in the transaction.

When the intermediate node receives the RREQ message, first, to avoid loops, check whether the same RREQ message has been introduced. If it is received, discard it. Otherwise, buffer the message and append its own neighbor node list, corresponding to the signal-to-noise ratio of the channel link, and update RREQ message and broadcast it. The payment cipher is managed by the bank. When the user opens an account, the bank registers and sends it to the user, and the key is generated by the algorithm chip, and the public key is stored in the bank's back-end server. When the bill is delivered to the bank, the bank enters the various elements of the bill on the counter terminal and transmits it to the verification machine through the bank network for verification. If the verification is correct, it indicates that the bill is indeed issued by the customer, and the bill elements are correct, so you can make payment. The public chain and alliance chain node operation agreement maintains the operation of the Internet of Things system. The alliance chain node generates new blocks at a time interval of 30 s. When the node cannot complete the block generation within 30 s, it will rotate to the next node.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the Social Science Foundation of Shaanxi Province of China under grant no. 2019D009 and the Shaanxi Province of China, Department of Education Project of Philosophy and Social Sciences Key Research Base under grant no. 19JZ052.

## References

- [1] J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang, and W. Zhao, "A survey on internet of things: architecture, enabling technologies, security and privacy, and applications," *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1125–1142, 2017.
- [2] X. Sun and N. Ansari, "EdgeIoT: mobile edge computing for the internet of things," *IEEE Communications Magazine*, vol. 54, no. 12, pp. 22–29, 2016.
- [3] Y. Yang, L. Wu, G. Yin, H. Zhao, and L. Li, "A survey on security and privacy issues in internet-of-things," *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1250–1258, 2017.
- [4] I. Yaqoob, E. Ahmed, I. A. T. Hashem et al., "Internet of things architecture: recent advances, taxonomy, requirements, and open challenges," *IEEE Wireless Communications*, vol. 24, no. 3, pp. 10–16, 2017.
- [5] O. Bello and S. Zeadally, "Intelligent device-to-device communication in the internet of things," *IEEE Systems Journal*, vol. 10, no. 3, pp. 1172–1182, 2016.
- [6] M. M. Zuberi and R. Levin, "Schumpeter's revenge: the gale of creative destruction: digital currencies and blockchain technology," *Banking & Financial Services Policy Report*, vol. 35, no. 5, pp. 1–8, 2016.
- [7] M. I. Mehar, C. L. Shier, A. E. Giambattista et al., "Understanding a revolutionary and flawed grand experiment in blockchain," *Journal of Cases on Information Technology*, vol. 21, no. 1, pp. 19–32, 2019.
- [8] D. S. Gong, "Financial transactions in ATM machines using speech signals," *International Journal of Engineering Research and Applications*, vol. 7, no. 1, pp. 25–28, 2017.
- [9] F. Balo, "Internet of things: a survey," *International Journal of Applied Mathematics Electronics and Computers*, vol. 4, no. 2016, pp. 104–110, 2016.
- [10] M. R. Palattella, M. Dohler, A. Grieco et al., "Internet of things in the 5G era: enablers, architecture, and business models," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 510–527, 2016.
- [11] A. V. Torsner and R. Buyya, "Fog computing: helping the internet of things realize its potential," *Computer*, vol. 49, no. 8, pp. 112–116, 2016.
- [12] M. A. Razzaque, M. Milojevic-Jevric, A. Palade, and S. Clarke, "Middleware for internet of things: a survey," *IEEE Internet of Things Journal*, vol. 3, no. 1, pp. 70–95, 2016.
- [13] J. Singh, T. Pasquier, J. Bacon, H. Ko, and D. Eyers, "Twenty security considerations for cloud-supported internet of things," *IEEE Internet of Things Journal*, vol. 3, no. 3, pp. 269–284, 2016.
- [14] A. Mosenia and N. K. Jha, "A comprehensive study of security of internet-of-things," *IEEE Transactions on Emerging Topics in Computing*, vol. 5, no. 4, pp. 586–602, 2017.
- [15] M. Amadeo, J. C. Campolo, A. D. MolinaroCorujo, R. L. Aguiar, and A. V. Vasilakos, "Information-centric networking for the internet of things: challenges and opportunities," *IEEE Network*, vol. 30, no. 2, pp. 92–100, 2016.
- [16] Y. Iera and J. Wen, "The IoT electric business model: using blockchain technology for the internet of things," *Peer-to-Peer Networking and Applications*, vol. 10, no. 4, pp. 983–994, 2017.
- [17] Z. Longchao, Y. X. Jianjun, and Y. Limei, "Research on congestion elimination method of circuit overload and transmission congestion in the internet of things," *Multi-media Tools and Applications*, vol. 76, no. 17, pp. 18047–18066, 2017.
- [18] J. W. Xue, X. K. Xu, and F. Zhang, "Big data dynamic compressive sensing system architecture and optimization algorithm for internet of things," *Discrete and Continuous Dynamical Systems-Series S*, vol. 8, no. 6, pp. 1401–1414, 2017.
- [19] H. S. Dhillon, H. Huang, and H. Viswanathan, "Wide-area wireless communication challenges for the internet of things," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 168–174, 2017.
- [20] A. Ouaddah, A. A. Abou Elkalam, and A. Ait Ouahman, "FairAccess: a new blockchain-based access control framework for the internet of things," *Security and Communication Networks*, vol. 9, no. 18, pp. 5943–5964, 2016.
- [21] M. Nitti, G. V. Piloni, and L. Atzori, "The virtual object as a major element of the internet of things: a survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1228–1240, 2016.
- [22] J. Ni, K. Zhang, X. Lin, and X. S. Shen, "Securing fog computing for internet of things applications: challenges and solutions," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 99, pp. 601–628, 2018.
- [23] D. Zhang, L. T. Yang, M. Chen, S. Zhao, M. Guo, and Y. Zhang, "Real-time locating systems using active RFID for internet of things," *IEEE Systems Journal*, vol. 10, no. 3, pp. 1226–1235, 2016.
- [24] K. Sood, Y. S. Yu, and Y. Xiang, "Software-defined wireless networking opportunities and challenges for internet-of-things: a review," *IEEE Internet of Things Journal*, vol. 3, no. 4, pp. 453–463, 2016.
- [25] S. P. Mohanty, U. Choppali, and E. Kougianos, "Everything you wanted to know about smart cities: the internet of things is the backbone," *IEEE Consumer Electronics Magazine*, vol. 5, no. 3, pp. 60–70, 2016.
- [26] J. Bernal Bernabe, A. F. Hernandez Ramos, and A. F. Skarmeta Gomez, "TACIoT: multidimensional trust-aware access control system for the internet of Things," *Soft Computing*, vol. 20, no. 5, pp. 1763–1779, 2016.

## Research Article

# Research on Lung Nodule Detection Based on Improved Target Detection Network

Ye Li <sup>1</sup>, Qian Wu <sup>1</sup>, Hongwei Sun <sup>1</sup> and Xuewei Wang <sup>1,2</sup>

<sup>1</sup>College of Computer Science and Engineering, Weifang University of Science and Technology, Weifang 262700, China

<sup>2</sup>Facility Horticulture Laboratory of Universities in Shandong, Weifang University of Science and Technology, Weifang 262700, China

Correspondence should be addressed to Hongwei Sun; [sunhongwei\\_doctor@wfust.edu.cn](mailto:sunhongwei_doctor@wfust.edu.cn)

Received 23 October 2020; Revised 27 November 2020; Accepted 2 December 2020; Published 17 December 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Ye Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Lung nodules are an early symptom of lung cancer. The earlier they are found, the more beneficial it is for treatment. However, in practice, Chinese doctors are likely to cause misdiagnosis. Therefore, deep learning is introduced, an improved target detection network is used, and public datasets are used to diagnose and identify lung nodules. This paper selects the Mask-RCNN network and uses the dense block structure of Densenet and the channel shuffle convolution method to improve the Mask-RCNN network. The experimental results prove that proposed algorithm is extremely effective.

## 1. Introduction

Pulmonary nodules [1, 2] are the result of the competition between unknown antigens and the body's cellular and humoral immune functions. It is very harmful to the human body. It is an early manifestation of lung cancer [3]. It usually appears as a circle in medical imaging. Shape or round shape, the lung tissue is complex, and it is difficult to distinguish lung nodules from blood vessels and bronchus in chest tissue very accurately based on the experience of clinicians and film readers. Vascular adhesion type and subpleural type are even more difficult in the screening of lung cancer.

In recent years, with the development of computer vision [4, 5] and artificial intelligence, the application of machine learning in medical image detection [6, 7] has also increased, among which machine learning is used in lung nodule detection the target detection network in deep learning [8, 9] can accurately locate the location of the region of interest and return its category. The common ones are R-CNN [10] series, SSD, and YOLO [11] series, among which R-CNN series mainly include R-CNN, fast R-CNN [12, 13], faster R-CNN, and Mask R-CNN [14, 15]. The advantage of the R-CNN series of target detection networks is that the

detection accuracy is high, but the disadvantage is that the detection time is long. The later SSD and YOLO series networks have fast detection speeds, but the accuracy is low; the accuracy of the YOLO-V3 network [16, 17] has been greatly improved. In terms of comprehensive detection speed and detection accuracy, YOLO-V3 network is often used.

In this paper, the YOLO-V3 target detection network in deep learning is selected and the network is improved. The reference network in the YOLO-V3 network is replaced with the SEnet network [18, 19], and the LIDC-IDRI public tuberculosis dataset [20] is used as the training dataset of the network. The trained network surpasses other target detection models and the unimproved YOLO-V3 network in many performances.

The rest of the paper is organized as follows. In Section 2, we introduce R-CNN network. In Section 3, an example is given to demonstrate the effectiveness of our method. In Section 4, the conclusions and future directions are given.

## 2. Network Improvement

This paper selects the mask R-CNN network, which has a higher accuracy rate in the field of medical imaging among



many models. The following introduces the Mask R-CNN network and its specific improvement methods.

*2.1. Introduction to the Dataset.* This paper selects the LIDC-IDRI public tuberculosis dataset as the training set of the network. The dataset is composed of chest medical image files (such as CT and X-ray film) and corresponding diagnosis result lesion labels. The data was collected by the National Cancer Institute (National Cancer Institute) in order to study early cancer detection in high-risk populations. This dataset contains a total of 1018 research examples. For the images in each example, four experienced thoracic radiologists performed a two-stage diagnosis and annotation. In the first stage, each physician independently diagnosed and marked the location of the patient. Three categories were marked: (1)  $\geq 3$  mm nodules, (2)  $< 3$  mm nodules, and (3)  $\geq 3$  mm nonnodules. In the subsequent second stage, each physician independently reviewed the labels of the other three physicians and gave their final diagnosis results. Such two-stage labeling can label all results as completely as possible while avoiding forced consensus. The image file is in Dicom format, which is a standard format for medical images. In addition to image pixels, there are some auxiliary metadata such as image type, image time, and other information. A CT image has  $512 \times 512$  pixels. Figure 1 is two randomly selected CT images.

*2.2. Mask R-CNN Network.* The detection of medical images pays more attention to the model performance. When the speed can meet the requirements, a network with higher detection accuracy should be selected as far as possible. The Mask R-CNN network is a highly accurate network, and its specific structure is shown in Figure 2.

Among them, CNN represents the benchmark network of the Mask R-CNN network, RPN represents the generation of the suggestion window network, ROIAlign represents the use of bilinear interpolation to obtain the region in the feature map corresponding to the ROI in the original image, the correspondence between the coordinates is preserved, and the mask branch represents FCN Internet.

It can be seen from the network structure that the mask R-CNN network finally outputs the results through two branches, the first branch outputs the background and object segmentation results, and the second branch outputs the classification and coordinate results. However, the benchmark network of the mask R-CNN network is the residual network, and it is not the best.

*2.3. Improve Mask R-CNN Network.* Densenet network is an improvement of residual network, which is a convolutional neural network with dense connections. In this network, there is a direct connection between any two layers, that is, the input of each layer of the network is the union of the outputs of all previous layers, and the feature map learned by this layer will also be directly passed to all subsequent layers are used as input. Figure 3 is a schematic diagram of Densenet's dense block. The structure of a block is as follows, which is the same as bottleneck in the residual.

The Densenet network is made up of dense blocks, and its specific structure is shown in Figure 4.

This paper selects the Densenet network as the reference network of the Mask R-CNN network, but the convolution method of the Densenet network will cause a lot of waste, and the experiment in this article is run on 3 GPUs, using the packet convolution method, in the packet volume. It is difficult to realize the information exchange between groups in the product. At the same time, the convolution method of the Densenet network will also cause a large amount of parameters. Therefore, this article uses the channel shuffle convolution method to reduce the amount of Densenet network parameters while also solving the grouping volume. There are many ways of product defects.

The convolution method in channel shuffle convolution is not the same as the convolution of the Densenet network. In the convolution of the Densenet network, a set of convolution kernels is responsible for a set of feature maps, while in the channel shuffle convolution, a convolution kernel is responsible for a feature map, which can greatly reduce the amount of parameters, but this will cause the loss of information between the same group of data. The shuffle operation can solve the problem of noncommunication of information in the group, and the shuffle operation can solve the problem of group and group convolution., the defect of not communicating information between groups. Figure 5 is a schematic diagram of channel shuffle, where input represents the input, GConv represents a grouped convolutional layer, Feature represents the feature map, and Output represents output.

Figure 5(a) represents a grouped convolution, and three colors represent three groups. It can be seen that there is no information exchange between each group of grouped convolution, and Figure 5(b) is the application of shuffle process, and it can be seen that there is an order to exchange information. Figure 5(c) is after shuffle is applied, and it can be seen that there are other groups of information between each group.

Change the convolution mode in the Densenet network to the channel shuffle convolution mode to get an improved network D-ShuffleNet network. In this paper, the D-ShuffleNet network is used as the reference network of the Mask R-CNN network to obtain the Per-T Mask R-CNN-II network.

To verify that the improvements made in this article are correct, four groups of networks are used to verify on the same small dataset. The four networks are Pre-T + Mask R-CNN-II improved by adding D-ShuffleNet network. The Pre + Mask R-CNN-II improved by the Densenet network, Mask R-CNN-II, and Mask R-CNN networks that only use the channel shuffle convolution method; the results are shown in Figure 6.

## 2.4. Network Training Strategy

*2.4.1. Activation Function.* Common activation functions include sigmoid, TANH, and Relu. The sigmoid function formula is as follows:

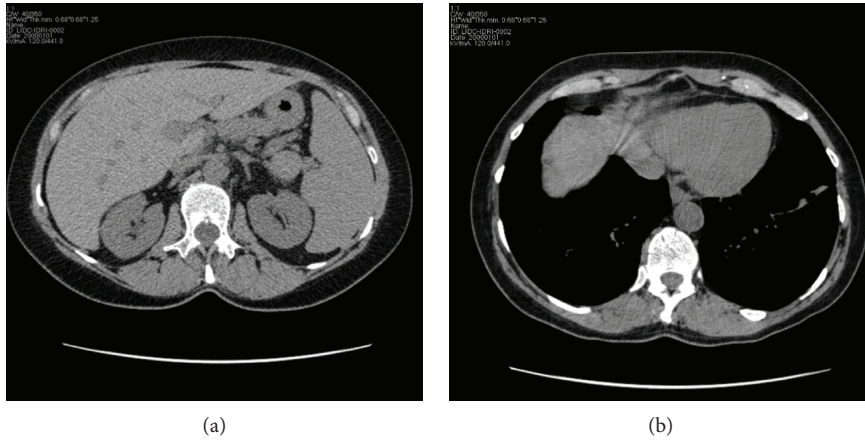


FIGURE 1: Picture of LIDC-IDRI data set.

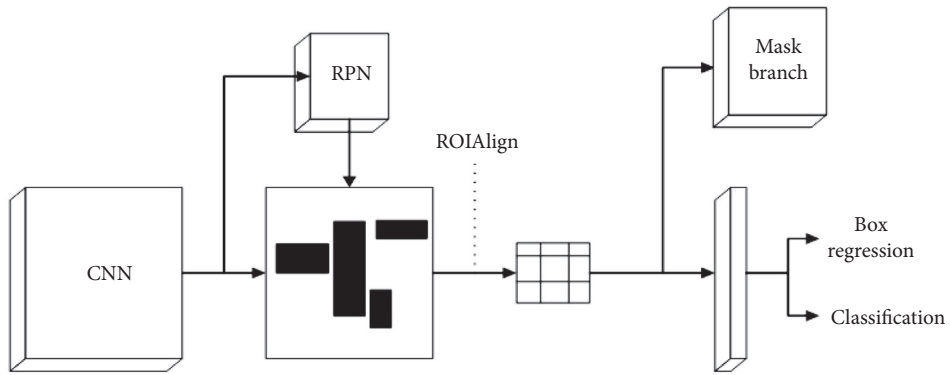


FIGURE 2: Mask R-CNN network structure.

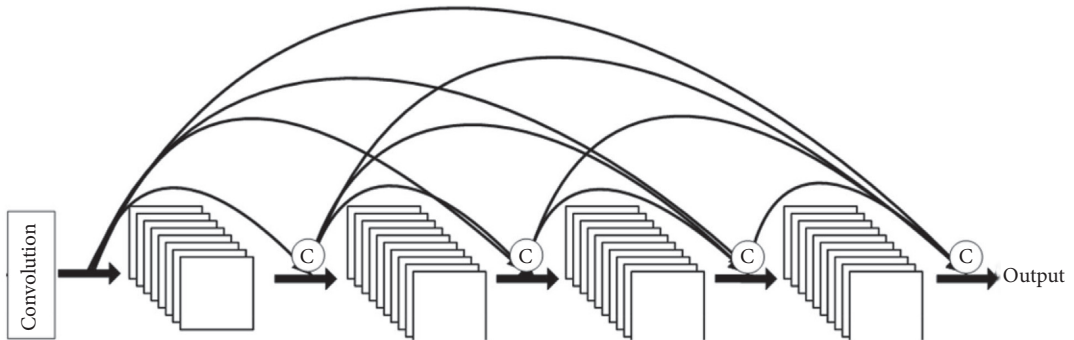


FIGURE 3: Schematic diagram of dense block in Densenet.

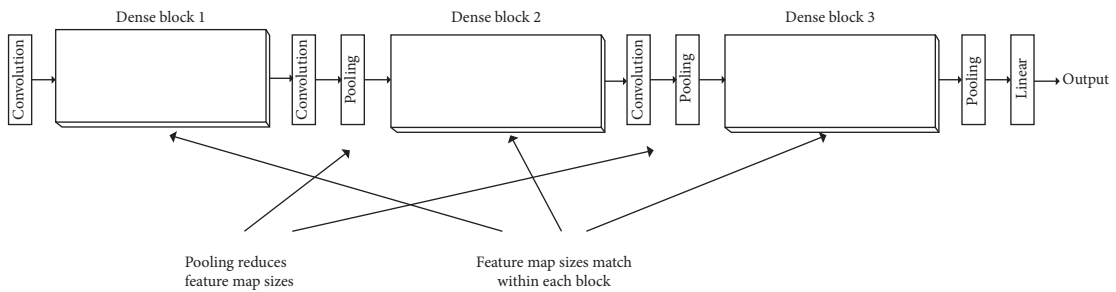


FIGURE 4: Schematic diagram of Densenet structure.

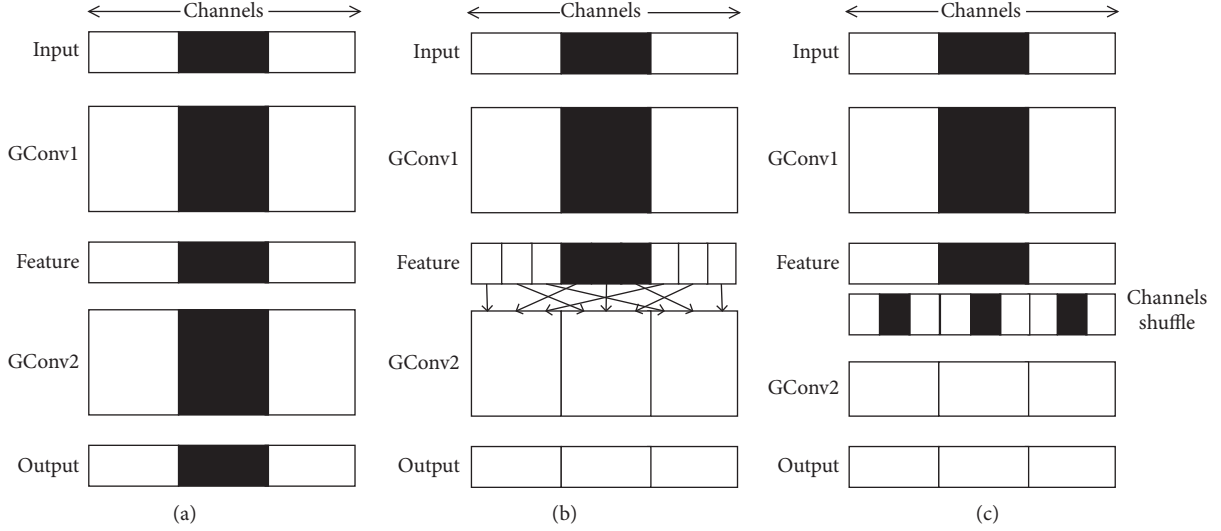


FIGURE 5: Schematic diagram of channel shuffle. (a) Grouped convolution. (b) Apply shuffle. (c) shuffle effect.

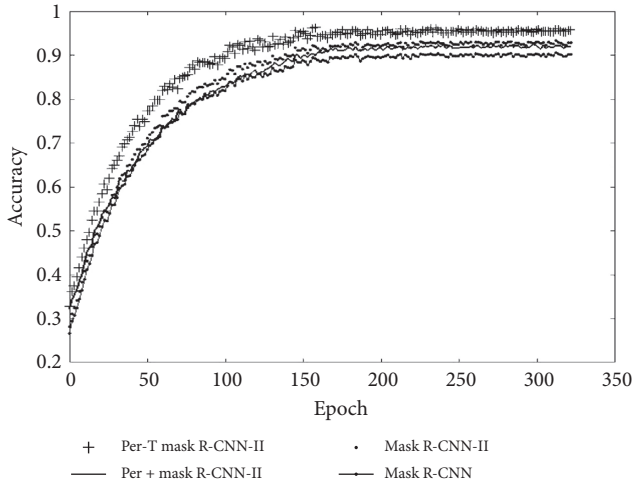


FIGURE 6: Schematic diagram of the four network results.

$$f(x) = \frac{1}{1 - e^{-x}}, \quad (1)$$

in which  $x$  represents the input and represents output. The output range from formula (1) is  $(0, 1)$ , which is not a 0-centralized distribution. During backpropagation, it is completely positive, so it will cause the weight parameter to update. The updated value is completely positive or completely negative. When the gradient is in the second or fourth quadrant, it will be difficult to find the optimal gradient. At the same time, since the gradient is close to 0, when the absolute value of the output value is large, the problem of gradient disappearance will be caused, and the sigmoid function will not be considered in general.

Equation (2) is the TANH function, which is better than the SIGMOID function:

$$f(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}}. \quad (2)$$

It can be concluded from the formula that the output range of the tanh function is  $(-1, 1)$ , which solves the problem that the sigmoid function cannot be distributed with 0 centralization, but it still does not solve the problem of vanishing gradient.

Equation (3) is the Relu function, which solves the problem

$$f(x) = \begin{cases} x, & x > 0, \\ 0, & x \leq 0. \end{cases} \quad (3)$$

It can be concluded that when the input is greater than 0, the gradient is always 1, which solves the problem of gradient disappearance. However, some parameters with input less than 0 are all dead, and the output is 0.

To solve the shortcomings of the Relu function, improvements were subsequently made. LReLU, PReLU, RReLU, and ELU appeared successively. The core idea of these four functions is to make the output not 0 when the output is less than 0 and then solve the parameter less than 0. For the death problem, this article uses the ELU transfer function:

$$f(x) = \begin{cases} \alpha(e^x - 1), & x < 0, \\ x, & x > 0. \end{cases} \quad (4)$$

**2.4.2. Learning Rate.** The initial value of the learning rate is usually set to 0.01, and it should be determined according to the actual situation. The usual practice is to set it to 0.01, and then initially iterate for about 10 epochs, generally checking the loss function and the transformation trend of accuracy if the loss can be reduced as well as increasing the accuracy rate; it means that the initial learning rate is generally

appropriate. We can try several times to choose a good initial learning rate.

The most common learning method is the step learning method. The learning rate is attenuated to one-tenth of the original every certain step size, which generally meets the requirements of a large learning rate in the early stage of training and a small learning rate in the later stage of training. As the gradient reaches a plateau, the training loss will be more difficult to improve. A saddle point is a point where the derivative of the function is zero but is not a local extremum on the axis. The difficulty of reducing loss comes from the saddle point, not the local minimum. If the training does not improve the loss, we can change the learning rate of each iteration according to some periodic function.

This article chooses the Warm Restart (Warm Restart) proposed by Loshchilov and Hutter and improves it accordingly. This method uses a cosine function as a periodic function and restarts the learning rate at the maximum value of each period. The improvement of this paper is to change the learning rate every certain step size so that the learning rate changes in a decreasing cosine manner. This improvement has a better application in the later stage of training.

**2.4.3. Loss Function and Regularization.** Loss functions generally include mean square error, maximum likelihood error, maximum posterior probability, cross-entropy loss function, cross-entropy function, and mean square error is an earlier loss function definition method, which measures the corresponding dimensions of the two distributions the sum of differences, the maximum likelihood error is from the perspective of probability, the model parameter theta that can perfectly fit the training example is solved, so that the probability  $p(y|x, \theta)$  is maximized and the posterior probability is maximized, that is, The maximum probability  $p(\theta|x, y)$  is actually equivalent to the maximum likelihood probability with a regularization term. It considers prior information and prevents “overfitting” by constraining the size of parameter values and the cross-entropy loss function measuring the similarity of two distributions  $p$  and  $q$ .

This paper chooses a relatively good cross-entropy loss function; formula (5) is the cross-entropy loss function formula:

$$C = -\frac{1}{n} \sum_x [y \ln a + (1 - y) \ln (1 - a)], \quad (5)$$

in which,  $y$  is the expected output and  $a$  is the actual output of the neuron.

Regularization is added in this article to prevent overfitting. Common regularizations include L1 regularization terms and L2 regularization term. In this paper, L2 regularization term is selected. Equation (6) is the objective function formed by adding L2 regularization term, which is the weight attenuation, and set to 0.9:

$$L = C + \frac{\lambda}{2n} \sum_w w^2. \quad (6)$$

**2.4.4. Optimizer.** Common optimizers include SGD, Momentum, Nesterov, RMSprop, and Adam. Among them, SGD is the earliest optimization method, but SGD is easy to converge to the local optimum, and in some cases may be trapped at the saddle point. Momentum can determine the relevant direction, accelerates SGD, suppresses oscillation, and speeds up the convergence, but it cannot improve the sensitivity adaptively. Nesterov and RMSprop also have certain shortcomings. The best optimization method now is the Adam optimization method. This article selects the Adam optimization method.

Adam is essentially RMSprop with a momentum term, which uses the first-order moment estimation and second-order moment estimation of the gradient to dynamically adjust the learning rate of each parameter. The main advantage of Adam is that, after bias correction, each iteration of the learning rate has a certain range, making the parameters more stable.

This article chooses the Adam optimizer, which uses the first-order moment estimation and the second-order moment estimation of the gradient to dynamically adjust the learning rate of each parameter. The main advantage of Adam is that, after bias correction, each iteration of the learning rate has a certain range, which makes the parameters relatively stable. The update formula is as follows:

$$\begin{aligned} m_t &= \beta_1 * m_{t-1} + (1 - \beta_1) * g_t, \\ v_t &= \beta_2 * v_{t-1} + (1 - \beta_2) * g_t^2, \\ \hat{m}_t &= \frac{m_t}{1 - \beta_1^t}, \\ \hat{v}_t &= \frac{v_t}{1 - \beta_2^t}, \\ \theta_{t+1} &= \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} * \hat{m}_t. \end{aligned} \quad (7)$$

### 3. Experimental Results

Practice is carried out using the Pytorch framework; training is carried out using the Adam optimizer and selecting the ELU function for the activation function. The convolution method in the Densenet network is changed to the channel shuffle convolution method, and the network D-ShuffleNet is improved.

This article conducted 7 sets of comparative experiments, namely, Pre-T+Mask R-CNN-II improved by adding D-ShuffleNet network and Pre+Mask R-CNN-II improved only by Densenet network, using only channel shuffle convolution Mask R-CNN-II, Mask R-CNN network, YOLO-V3 network, Fast R-CNN network, and SSD network, it uses the same dataset for training. After training, the

TABLE 1: Confusion matrix of classification results.

| The true situation   | Forecast result      |                            |
|----------------------|----------------------|----------------------------|
|                      | Positive example (P) | Counterexample (N)         |
| Positive example (P) | TP (real case)       | FN (false counterexample)  |
| Counterexample (N)   | FP (false positive)  | TN (true negative example) |

experimental results of 7 sets of experiments are obtained. In this paper, mAP and ROC curves are used as evaluation criteria to analyze and evaluate the network results. AP in mAP refers to the area of the P-R curve of each class, mAP is the average of all APs, P in the P-R curve refers to precision, which is accuracy, and R refers to recall, which is the recall rate. The above concepts can be derived from the concept of confusion matrix. The confusion matrix table of classification results is shown in Table 1.

The formulas of recall rate and accuracy rate are shown in equations (8) and (9):

$$R = \frac{TP}{TP + FN}, \quad (8)$$

$$P = \frac{TP}{TP + FP}. \quad (9)$$

Table 2 shows the specific values of the seven groups of models after training, including the AP value and mAP value of each category.

The order of mAP size in Table 2 is Pre-T + Mask R-CNN-II, Pre + Mask R-CNN-II, Mask R-CNN-II, Mask R-CNN, YOLO-V3, Faster R-CNN, and SSD.

It can be seen from Table 2 that the training effect of the Pre-T + Mask R-CNN-II model is the best, indicating that the network proposed in this paper is suitable for lung nodule target detection.

Regarding model evaluation criteria, in addition to the most commonly used ones, it is often necessary to evaluate the model through the ROC curve. ROC space defines the false positive rate (False Positive Rate, FPR) as the X-axis, and the true positive rate (True Positive Rate) Rate, TPR for short) is defined as the Y-axis. These two values are calculated from the four values in Table 1.

For TPR, in all samples that are actually positive, the ratio of correctly judged positive is

$$TPR = \frac{TP}{TP + FN}. \quad (10)$$

For FPR, among all samples that are actually negative examples, the ratio of falsely judged positive examples is

$$FPR = \frac{FP}{FP + TN}. \quad (11)$$

The ROC curve can reflect the performance of the model well, and its area is AUC. The larger the AUC value, the better the performance of the model. Figure 7 is the ROC curve.

After calculation, the area of Pre-T + Mask R-CNN-II, Pre + Mask R-CNN-II, Mask R-CNN-II, Mask R-CNN, YOLO-V3, Faster R-CNN, and SSD is 0.942, 0.935, 0.916,

TABLE 2: Results of each model.

|                       | mAP   |
|-----------------------|-------|
| Pre-T + Mask R-CNN-II | 0.927 |
| Pre + Mask R-CNN-II   | 0.921 |
| Mask R-CNN-II         | 0.907 |
| Mask R-CNN            | 0.893 |
| YOLO-V3               | 0.881 |
| Faster R-CNN          | 0.871 |
| SSD                   | 0.865 |

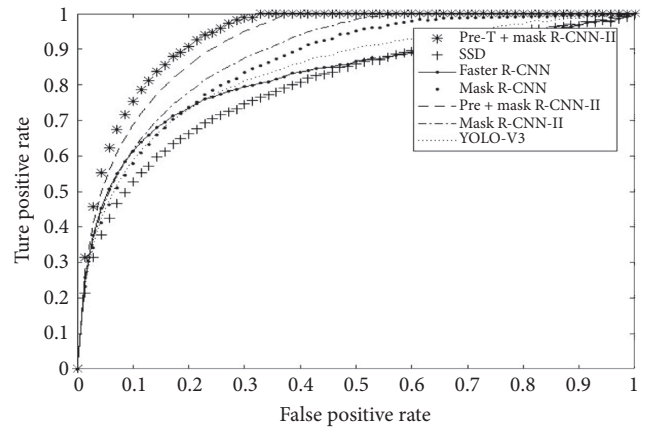


FIGURE 7: ROC curve.

0.902, 0.893, 0.882, and 0.877, which are consistent with the results obtained by the mAP evaluation system.

## 4. Conclusions

This paper proposes a new network D-ShuffleNet network, by combining the Densenet network and channel shuffle convolution method, and then proposes a new target detection network. Through the last seven sets of comparative experiments, it is proved that the network proposed in this paper has better performance than other networks. All are good, but the network still has room for improvement. The next step is to improve the network performance and improve the recognition accuracy.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments


This study was supported by the Facility Horticulture Laboratory of Universities in Shandong with project nos. 2019YY003, 2018RC002, 2018YY016, 2018YY043, and 2018YY044; Soft Science Research Project of Shandong Province with project no. 2019RKA07012; Research and Development Plan of Applied Technology in Shouguang with project no. 2018JH12; 2018 Innovation Fund of Science and Technology Development Centre of the China Ministry of Education in project no. 2018A02013; 2019 Basic Capacity Construction Project of Private Colleges and Universities in Shandong Province; and Weifang Science and Technology Development Programme with project no. 2019GX081 and 2019GX082.

## References

- [1] S. Wang, S. Wang, S. Zhang, and Y. Wang, "Time-ResNeXt for epilepsy recognition based on EEG signals in wireless networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2020, p. 195, 2020.
- [2] Z. Chen and H. Ahn, "Item response theory based ensemble in machine learning," *International Journal of Automation and Computing*, vol. 17, no. 5, pp. 621–636, 2020.
- [3] S. Wang, G. He, S. Wang, S. Zhang, and F. Fan, "Research on recognition of medical image detection based on neural network," *IEEE Access*, vol. 8, pp. 94947–94955, 2020.
- [4] R. O. Lloyd, J. M. O'Toole, E. Pavlidis, P. M. Filan, and G. B. Boylan, "Electrographic seizures during the early postnatal period in preterm infants," *The Journal of Pediatrics*, vol. 187, pp. 18–25, 2017.
- [5] D. M. Murray, G. B. Boylan, I. Ali, C. A. Ryan, B. P. Murphy, and S. Connolly, "Defining the gap between electrographic seizure burden, clinical expression and staff recognition of neonatal seizures," *Archives of Disease in Childhood—Fetal and Neonatal Edition*, vol. 93, no. 3, pp. F187–F191, 2008.
- [6] American Epilepsy Society, *Facts and Figures*, Chicago, IL, USA, 2020, [https://www.aesnet.org/for\\_patients/facts\\_figures](https://www.aesnet.org/for_patients/facts_figures).
- [7] Z. Huang, X. Xu, H. Zhu, and M.C. Zhou, "An efficient group recommendation model with multiattention-based neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 11, pp. 4461–4474, 2020.
- [8] T. N. Tsuchida, C. J. Wusthoff, R. A. Shellhaas et al., "American Clinical Neurophysiology Society standardized EEG terminology and categorization for the description of continuous EEG monitoring in neonates: report of the American Clinical Neurophysiology Society critical care monitoring committee," *Journal of Clinical Neurophysiology*, vol. 30, no. 2, pp. 161–173, 2013.
- [9] P. Srinivasakumar, J. Zempel, S. Trivedi et al., "Treating EEG seizures in hypoxic ischemic encephalopathy: a randomized controlled trial," *Pediatrics*, vol. 136, no. 5, pp. e1302–e1309, 2015.
- [10] A. Krumholz, S. Wiebe, G. Gronseth et al., "Practice parameter: evaluating an apparent unprovoked first seizure in adults (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology and the American Epilepsy Society," *Neurology*, vol. 69, no. 21, 2007.
- [11] V. Srinivasan, C. Eswaran, and N. Sriraam, "Approximate entropy-based epileptic EEG detection using artificial neural networks," *IEEE Transactions on Information Technology in Biomedicine*, vol. 11, no. 3, pp. 288–295, 2007.
- [12] L. Guo, D. Rivero, J. Dorado, J. R. Rabuñal, and A. Pazos, "Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks," *Journal of Neuroscience Methods*, vol. 191, no. 1, pp. 101–109, 2010.
- [13] R. Tetzlaff, C. Niederhöfer, and P. Fischer, "Automated detection of a pre seizure state: non-linear EEG analysis in epilepsy by cellular nonlinear networks and Volterra systems," *International Journal of Circuit Theory and Applications*, vol. 34, no. 1, pp. 89–108, 2006.
- [14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, <https://arxiv.org/abs/1409.1556>.
- [15] H. Liang, J. Zou, Z. Li, M. J. Khan, and Y. Lu, "Dynamic evaluation of drilling leakage risk based on fuzzy theory and PSO-SVR algorithm," *Future Generation Computer Systems*, vol. 95, pp. 454–466, 2019.
- [16] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, June 2015.
- [17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016.
- [18] H. Liang, J. Zou, K. Zuo, and M. J. Khan, "An improved genetic algorithm optimization fuzzy controller applied to the wellhead back pressure control system," *Mechanical Systems and Signal Processing*, vol. 142, Article ID 106708, 2020.
- [19] Y. Zhang, R. Zhu, Z. Chen, J. Gao, and D. Xia, "Evaluating and selecting features via information theoretic lower bounds of feature inner correlations for high-dimensional data," *European Journal of Operational Research*, vol. 63, 2020 In press.
- [20] H. Liang, A. Xian, M. Mao, P. Ni, and H. Wu, "A research on remote fracturing monitoring and decision-making method supporting smart city," *Sustainable Cities and Society*, vol. 62, Article ID 102414, 2020.

## Research Article

# Robust Adaptive Control for a Class of T-S Fuzzy Nonlinear Systems with Discontinuous Multiple Uncertainties and Abruptly Changing Actuator Faults

Xin Ning,<sup>1,2</sup> Yao Zhang,<sup>1,2</sup> and Zheng Wang<sup>3,4,2</sup> 

<sup>1</sup>School of Astronautics, Northwestern Polytechnical University, Xi'an 710072, China

<sup>2</sup>National Key Laboratory of Aerospace Flight Dynamics, Northwestern Polytechnical University, Xi'an 710072, China

<sup>3</sup>Research Center for Unmanned System Strategy Development, Northwestern Polytechnical University, Xi'an 710072, China

<sup>4</sup>Unmanned System Research Institute, Northwestern Polytechnical University, Xi'an 710072, China

Correspondence should be addressed to Zheng Wang; wz\_nwpu@126.com

Received 15 September 2020; Revised 21 November 2020; Accepted 2 December 2020; Published 16 December 2020

Academic Editor: Paolo Bellavista

Copyright © 2020 Xin Ning et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the complex environment, the suddenly changing structural parameters and abrupt actuator failures are often encountered, and the negligence or improper handling method may induce undesired or unacceptable results. In this paper, taking the suddenly changing structural parameters and abrupt actuator failures into consideration, we focus on the robust adaptive control design for a class of heterogeneous Takagi–Sugeno (T-S) fuzzy nonlinear systems subjected to discontinuous multiple uncertainties. The key point is that the switch modes not only vary with the system time but also vary with the system states, and the intrinsic heterogeneous characteristics make it difficult to design stable controllers. Firstly, the concepts of differential inclusion are introduced to describe the heterogeneous fuzzy systems. Meanwhile, a fundamental lemma is provided to demonstrate the criteria of the boundness for a Filippov solution. Then, by using the set-valued Lie derivative of the Lyapunov function and introducing a vector of specific continuous functions, the closed-loop T-S fuzzy differential inclusion systems are proved to be ultimately bounded. The sufficient conditions for system stability are derived in term of linear matrix inequalities (LMIs), which can be solved directly. Finally, a numerical example is provided to illustrate the effectiveness of the proposed control algorithm.

## 1. Introduction

As it is well-known, the T-S fuzzy model is a powerful tool for the analysis and control design of nonlinear systems [1–4]. Therefore, a great wealthy of results has been achieved for T-S fuzzy systems in the past decades. In [5], the parameterized linear matrix inequality technique has been investigated for T-S fuzzy control systems. In [6], by using the LMIs and sum-of-squares-based approach, an output regulator has been constructed for the polynomial fuzzy control systems. In [7, 8], two fuzzy sliding mode control methods have been developed for the T-S fuzzy systems suffering from both matched and unmatched uncertainties. The fault-tolerant control approaches for T-S fuzzy nonlinear systems have been investigated in [9, 10]. As a further

development, a finite-time fault-tolerant control structure of T-S fuzzy nonlinear systems has been synthesized in [11]. The event-triggered control structures for T-S fuzzy nonlinear systems can be found in [12, 13]. In [14, 15], the robust filters have been constructed for continuous and discrete-time T-S fuzzy systems, respectively. For the delayed T-S fuzzy systems with and without stochastic perturbations, the stability analysis and stabilization methods have been provided in [16, 17]. Chang et al. [18] focused on the robust adaptive control design for a class of heterogeneous T-S fuzzy nonlinear systems subjected to discontinuous multiple uncertainties.

It is also well-known that the uncertainties and disturbances are often encountered in many practical systems. In [19, 20], the active disturbance rejection control methods

have been reported for the nonlinear systems with uncertainties. In [21–24], several antidisturbance controllers have been synthesized by constructing the disturbance observers. The composite antidisturbance controllers can be found in [25–29]. Aiming at the uncertainties existing in the quantized control systems, several robust and adaptive controllers have been proposed in [30–32]. For the unknown nonlinearities existing in the control systems, fruitful results have been reported. In [33], a fuzzy adaptive output feedback controller has been proposed for the multi-input and multioutput nonlinear systems with completely unknown nonlinear functions. For a class of switched stochastic nonlinear systems in a pure-feedback form, a fuzzy observer is constructed to approximate unmeasurable states in [34, 35]. In [36], an adaptive fuzzy tracking control problem has been investigated for a class of nonstrict-feedback systems with unmeasured states and unknown nonlinearities. For the state constrained control systems with unknown nonlinear functions, two adaptive control results have been reported in [37, 38]. In [39], the problem of adaptive neural finite-time tracking control for uncertain nonstrict-feedback nonlinear systems with input saturation has been studied.

In spite of the progress, in the aforementioned control results, a vital problem on heterogeneous uncertainties, in the sense that the parameters and the structures of the uncertainties keep switching with both the system time and the system states, was omitted. This kind of uncertainties and system nonlinearities are often encountered in many physical systems [40,41], and it is of significant importance to develop an effective controller for the heterogeneous uncertain nonlinear systems. When system-state-based switching uncertainties and system structures are taken into consideration, the traditional analysis and control methods become invalid. Furthermore, the control approaches of conventional switched systems cannot be applied neither because most of the switched systems are required to be

purely time-based [42–44]. Moreover, when the fuzzy modeling methods and fuzzy control strategies are introduced in the heterogeneous uncertain system, the control design problem becomes more challenging and interesting. As far as the authors know, no results have been reported for the T-S fuzzy systems subjected to heterogeneous uncertainties and system structures. For the purpose of improving the practicability of the proposed algorithm, the abruptly changing actuator faults are also considered. Motivated by the above considerations, this paper is committed to develop an effective control structure for the T-S fuzzy heterogeneous systems with discontinuous multiple uncertainties and abruptly changing actuator faults. Compared to the existing literature, the main contributions of this paper are as follows:

- (i) To the best of the authors' knowledge, it is the first control solution for the T-S fuzzy systems subjected to heterogeneous uncertainties and system structures, which keep switching with both the system time and the system states.
- (ii) A fundamental lemma is provided to demonstrate the criteria of the boundness for a Filippov solution, establishing the mathematical fundamentals for the adaptive control of differential inclusion systems.
- (iii) By proposing a specific vector of continuous functions, the closed-loop multivariable T-S fuzzy differential inclusion systems are proved to be ultimately bounded for the first time.

## 2. Problem Formulation and Preliminaries

**2.1. Problem Statement.** Consider the following T-S fuzzy nonlinear systems.

**Plant Rule:** IF  $\theta_{i,1}(t)$  is  $\mu_{i,1}$ ,  $\theta_{i,2}(t)$  is  $\mu_{i,2}$ ,  $\dots$  and  $\theta_{i,p}(t)$  is  $\mu_{i,p}$ , THEN

$$\dot{x}(t) = \begin{cases} f_i^1(t, x(t)) + A_i^1 x + B_i u, & \text{if } (t, x(t)) \in G_1, \\ f_i^2(t, x(t)) + A_i^2 x + \Delta A_i(t)x + B_i(u + d(t)), & \text{if } (t, x(t)) \in G_2, \\ f_i^3(t, x(t)) + A_i^3 x + \Delta A_i(t)x + B_i(\Lambda(t)u + \varepsilon_u(t) + d(t)), & \text{if } (t, x(t)) \in G_3, \end{cases} \quad (1)$$

where  $u(t) \in \mathbb{R}^m$  is the input signal of the system and  $x(t) \in \mathbb{R}^n$  are the system state vector.  $\theta(t) = [\theta_1(t), \theta_2(t), \dots, \theta_p(t)]$  are the premise variables.  $\mu_{i,1}, \mu_{i,2}, \dots, \mu_{i,p}$ ,  $i \in Y = \{1, 2, \dots, r\}$  are the fuzzy sets.  $r$  is the number of IF-THEN rules.  $G_k$ ,  $k = 1, 2, 3$  are the open connected sets satisfying that  $\cup_{k=1}^3 G_k = \mathbb{R}_+ \times \mathbb{R}^n$  and  $G_{k_1} \cap G_{k_2} = \emptyset$ ,  $\forall k_1, k_2 \in \{1, 2, 3\}, k_1 \neq k_2$ .  $S = \cup_{k=1}^3 \partial G_k$  is of Lebesgue measure zero, where  $\partial \Omega$  represents the boundary set of a set  $\Omega$ . For  $\forall i \in Y$ ,  $k = 1, 2, 3$ ,  $A_i^k \in \mathbb{R}^{n \times n}$ ,  $B_i \in \mathbb{R}^{n \times m}$ , and  $D \in \mathbb{R}^{n \times p_3}$  are all known matrices.  $f_i^k(t, x(t))$  is a vector of unknown nonlinear functions.  $\Delta A_i(t) \in \mathbb{R}^{n \times n}$  is an unknown matrix varying with the time.  $\Lambda(t)$  is a time-varying diagonal matrix of remanent actuator effectiveness and  $\varepsilon_u(t)$  is a vector of

actuator deviations when the actuator failure occurs.  $d(t)$  represents the external disturbance. The symbols used in the paper can be found in Table 1.

*Remark 1.* It should be highlighted that system model (1) can reflect the actual situation of many practical engineering systems and possesses important research significance. If  $(t, x(t))$  stays in  $G_1$ , the system is under a normal condition. When  $(t, x(t))$  enters  $G_2$  from  $G_1$ , the system uncertainties  $\Delta A_i(t)$  suddenly appear and external disturbances  $d(t)$  grow rapidly. Finally, if the system states travel into  $G_3$ , the actuator faults abruptly come out. Accompanied by  $(t, x(t))$  entering into different regions, the system matrix and the nonlinear function vector will change.



TABLE 1: The symbols used in the paper.

| Variable           | Implication                               |
|--------------------|---|
| $\theta(t)$        | Premise variables                         |
| $G_k$              | Open connected sets                       |
| $\partial G_k$     | Lebesgue measure zero                     |
| $\partial\Omega$   | Boundary set of a set $\Omega$            |
| $f_i^k(t, x(t))$   | Unknown nonlinear functions               |
| $A_i^k, B_i, D$    | Known matrices                            |
| $\Delta A_i(t)$    | Unknown matrix varying with time          |
| $\Lambda(t)$       | Matrix of remanent actuator effectiveness |
| $\varepsilon_u(t)$ | Vector of actuator deviations             |
| $d(t)$             | External disturbance                      |

It is supposed that  $\theta_i(t)$  never depends on  $u(t)$ ,  $\varepsilon(t)$ , and  $d(t)$ . Define

$$h_i(\theta(t)) = \frac{\prod_{j=1}^p \mu_{i,j}(\theta_j(t))}{\sum_{i=1}^r \prod_{j=1}^p \mu_{i,j}(\theta_j(t))}. \quad (2)$$

$\forall i \in \Upsilon, j = 1, 2, \dots, p, \mu_{i,j}(\theta_j(t))$  denote the degree of membership  $\theta_j(t)$  in  $\mu_{i,j}$ .  $h_i(\theta(t))$  is the fuzzy basis function. Clearly, for any  $i \in \Upsilon$ ,  $h_i(\theta(t)) \geq 0$  and  $\sum_{i=1}^r h_i(\theta(t)) = 1$ . Therefore, the dynamics of system (1) can be rewritten as follows:

$$\dot{x}(t) = g(t, x(t)),$$

$$g(t, x(t)) = \begin{cases} \sum_{i=1}^r h_i(\theta(t)) g_i^1(t, x(t)) & \text{if } (t, x(t)) \in G_1, \\ \sum_{i=1}^r h_i(\theta(t)) g_i^2(t, x(t)) & \text{if } (t, x(t)) \in G_2, \\ \sum_{i=1}^r h_i(\theta(t)) g_i^3(t, x(t)) & \text{if } (t, x(t)) \in G_3, \end{cases} \quad (3)$$

where

$$\begin{aligned} g_i^1(t, x(t)) &= f_i^1(t, x(t)) + A_i^1 x + B_i u, \\ g_i^2(t, x(t)) &= f_i^2(t, x(t)) + A_i^2 x + \Delta A_i(t) x + B_i(u + d(t)), \\ g_i^3(t, x(t)) &= f_i^3(t, x(t)) + A_i^3 x + \Delta A_i(t) x + B_i(\Lambda(t) u \\ &\quad + \varepsilon_u(t) + d(t)). \end{aligned} \quad (4)$$

Our control objective is to design an adaptive controller such that system states  $x(t)$  can converge into a desired compact set in the presence of the discontinuous multiple uncertainties and abruptly changing actuator faults.

To achieve the control objective, the following assumptions are necessary.

*Assumption 1.* For any  $i \in \Upsilon, k = 1, 2, 3, f_i^k(t, x(t))$  satisfy  $f_i^k(0, x(0)) = 0$  and

$$\|f_i^k(t, x_1(t)) - f_i^k(t, x_2(t))\| \leq \|U_i^k(x_1(t) - x_2(t))\|. \quad (5)$$

*Assumption 2.* For any  $i \in \Upsilon$ ,

$$\Delta A_i(t) = M_i \Xi_i(t) N_i, \quad (6)$$

where  $\Xi_i(t)$  are unknown time-varying matrices satisfying  $\Xi_i^T(t) \Xi_i(t) \leq I$  and  $M_i$  and  $N_i$  are known matrices.

*Assumption 3.* The discontinuous disturbances  $d_i(t)$  are Lebesgue measurable and locally bounded, i.e.,  $\|d_i(t)\| \leq D_i$ , where  $D_i$  is an unknown positive constant.

*Assumption 4.* It is supposed that  $\Lambda(t)$  and  $\varepsilon_u(t)$  are both Lebesgue measurable, and there exist lower and upper bounds for  $\Lambda(t)$  and  $\varepsilon_u(t)$ , respectively. In other words,  $\lambda_{\min}[\Lambda(t)] > \varepsilon_\Lambda > 0, \|\varepsilon_u(t)\| \leq D_u$ .  $\varepsilon_\Lambda$  and  $D_u$  are positive constants.

*2.2. Preliminaries.* Consider a nonlinear system:

$$\dot{x} = g(x(t)), \quad (7)$$

where  $x \in \mathbb{R}, g: \mathbb{R} \times [0, \infty) \rightarrow \mathbb{R}$  is Lebesgue measurable, essentially locally bounded and uniformly in  $t$ . Moreover, there exist discontinuities in  $g(x(t))$ .

*Definition 1.* A vector  $x(t)$  is called a Filippov solution of differential equation (7) over  $[0, \infty)$  if  $x(t)$  is absolutely continuous, and for almost everywhere  $t \in [0, \infty)$ ,

$$\dot{x} \in F[g](x(t)), \quad (8)$$

where  $\mathcal{F}[g](x(t))$  is a upper semicontinuous set-valued map defined by

$$F[g](x(t)) \stackrel{\Delta}{=} \bigcap_{\rho > 0} \bigcap_{\mu(N)=0} \bar{co} g\left(\frac{B(x(t), \rho)}{N}\right), \quad (9)$$

where  $\bigcap_{\mu(N)=0}$  denotes the intersection over sets of Lebesgue measure zero,  $\bar{co}$  represents the convex closure, and  $B(x(t), \rho) = \{z \in \mathbb{R}^n \mid \|x(t) - z\| < \rho\}$

*Definition 2* (see [45]). Given a locally Lipschitz function  $V(x(t))$ , the generalized gradient of  $V(x(t))$  is defined by

$$\partial V(x(t)) = \bar{co} \{\lim \nabla V(x^*(t)) \mid x^*(t) \rightarrow x(t), x^*(t) \notin \Omega_v\}, \quad (10)$$

where  $\Omega_v$  is the set of measure zero and  $\nabla V$  is not defined. Moreover, the set-valued Lie derivative of a  $V(x(t))$  is defined as

$$DV(x(t)) \stackrel{\Delta}{=} \bigcap_{\xi \in \partial V(x(t))} \xi^T F[g](x(t)). \quad (11)$$

*Definition 3.* In this paper, the generalized sections for variables, vectors, and matrices are defined. For  $a, b \in \mathbb{R}, a < b$ , define  $[b, a] = [a, b] = [\min\{a, b\}, \max\{a, b\}]$ . For  $a, b \in \mathbb{R}^n$ , define  $[a, b] = [b, a] = ([a_i, b_i])_n$ , where  $a_i$  is the  $i$ th component of  $a$ . For  $A, B \in \mathbb{R}^{m \times n}$ , define  $[A, B] = [B, A] = ([A_{ij}, B_{ij}])_{m \times n}$ .

**Lemma 1.** Consider nonlinear system (7). Suppose  $g(\cdot)$  is Lebesgue measurable and  $x \rightarrow g(x(t))$  is bounded. Let

$V: \mathbb{R} \times [0, \infty) \rightarrow \mathbb{R}$  be locally Lipschitz and regular such that

$$\begin{aligned} \alpha_1(x(t)) \leq V(x(t)) \leq \alpha_2(x(t)), \quad \forall t \geq 0, \\ DV(x(t)) \leq -\gamma_1 V(x(t)) + \gamma_2, \end{aligned} \quad (12)$$

where  $\alpha_1(\cdot)$  and  $\alpha_2(\cdot)$  are  $K_\infty$  functions,  $\gamma_1 > 0$  and  $\gamma_2 > 0$  are constants,  $x(t)$  is a Filippov solution of (7) with initial value  $x(t_0) \in \Omega_2 = \{x \in \mathbb{R}^n | \|x\| < r, \alpha_2(x) \leq \rho\}$ , and  $r > 0$  and  $\rho > 0$  are constants satisfying that  $\gamma_2/\gamma_1 < \rho < \min_{\|x(t)\|=r} \alpha_1(x(t))$ . Then,  $x(t)$  is bounded and converges to a compact set:

$$\Omega_f = \left\{ x \in \mathbb{R}^n | V(x) \leq \frac{\gamma_2}{\gamma_1} \right\}. \quad (13)$$

*Proof.* Define  $\Omega_V = \{x \in \mathbb{R}^n | \|x\| < r, V(x) \leq \rho\}$ . Since  $\alpha_2(x) \leq \rho \implies V(x) \leq \rho$ , it is obvious that  $\Omega_2 \subset \Omega_V$ . On the contrary, with the aid of  $V(x) \leq \rho \implies \alpha_1(x) \leq \rho$ , we can easily get that  $\Omega_V \subset \Omega_1$ . Moreover, considering  $\rho < \min_{\|x(t)\|=r} \alpha_1(x(t))$ , we know that  $\Omega_1 = \{x \in \mathbb{R}^n | \alpha_1(x) \leq \rho\}$  is an interior set of  $B_r = \{x \in \mathbb{R}^n | \|x\| < r\}$ . Hence,

$$\Omega_2 \subset \Omega_V \subset \Omega_1 \subset B_r. \quad (14)$$

If  $V(x(t)) \leq \gamma_2/\gamma_1$ , it is clear that  $x(t)$  stays in  $\Omega_f$ , which is  $x(t) \in \Omega_f \subset \Omega_V \subset B_r$ . On the contrary, for any  $x(t) \notin \Omega_f$ , we can obtain that  $V(x(t)) > \gamma_2/\gamma_1$  and  $\dot{V} < a.e. 0$ . Hence,  $V(x(t))$  is nonincreasing. Furthermore, for any  $x(t_0) \in \Omega_2$ , the solution of (7) stays in  $\Omega_V$ , which is  $x(t) \in \Omega_V \subset B_r$ . Therefore, it can be concluded that the solution of system (7) is bounded for all  $t \geq t_0$ . Furthermore, it follows from (12) that

$$\begin{aligned} \int_{t_0}^t [-\gamma_1 V(x(\tau)) + \gamma_2] d\tau \leq \int_{t_0}^t \dot{V}(x(\tau)) d\tau = V(x(t_0)) \\ - V(x(t)) \leq V(x(t_0)). \end{aligned} \quad (15)$$

Therefore,  $\int_{t_0}^t [-\gamma_1 V(x(\tau)) + \gamma_2] d\tau$  is bounded for all  $t \geq t_0$ . It should be noted that the existence of  $\int_{t_0}^t [-\gamma_1 V(x(\tau)) + \gamma_2] d\tau$  can be guaranteed because  $-\gamma_1 V(x(t)) + \gamma_2$  is locally Lipschitz and regular by definition. According to Barbalat's Lemma [46], it can be obtained that  $-\gamma_1 V(x(t)) + \gamma_2 \rightarrow 0$  as  $t \rightarrow \infty$ . Therefore, we know that  $x(t)$  will converge into  $\Omega_f$  finally. The proof is complete.  $\square$

**Lemma 2.** Given any constant  $\varepsilon > 0$  and any vector  $\xi \in \mathbb{R}^n$ , the following inequality holds:

$$\|\xi\| < \frac{\xi^T \xi}{\sqrt{\xi^T \xi + \varepsilon^2}} + \varepsilon. \quad (16)$$

*Proof.* Since  $\varepsilon > 0$  and  $\sqrt{\xi^T \xi + \varepsilon^2} > 0$ , it can be easily get that

$$\begin{aligned} \left[ \xi^T \xi + \varepsilon \sqrt{\xi^T \xi + \varepsilon^2} \right]^2 - \left[ \|\xi\| \sqrt{\xi^T \xi + \varepsilon^2} \right]^2 \\ = 2\varepsilon \xi^T \xi \sqrt{\xi^T \xi + \varepsilon^2} + \varepsilon^4 > 0. \end{aligned} \quad (17)$$

Hence, we know that

$$\|\xi\| \sqrt{\xi^T \xi + \varepsilon^2} < \xi^T \xi + \varepsilon \sqrt{\xi^T \xi + \varepsilon^2}. \quad (18)$$

By dividing  $\sqrt{\xi^T \xi + \varepsilon^2}$  in both sides of (18), inequality (16) can be obtained. The proof is completed.  $\square$

### 3. Main Results

**3.1. Control Design.** In the following text, the robust adaptive control problem for the concerned T-S fuzzy discontinuous nonlinear systems will be addressed.

In view of (3), according to Definition 1, we can get the following differential inclusion:

$$\begin{aligned} \dot{x} \in \sum_{i=1}^r h_i(\theta(t)) \{ F[f_i](t, x(t)) + F[A_i]x(t) \\ + F[\Delta \bar{A}_i(t)]x(t) + B_i F[\bar{\Lambda}(t)]u(t) + B_i F[\bar{d}(t)] \}, \end{aligned} \quad (19)$$

where

$$\begin{aligned} f_i &= f_i^k(t, x(t)) \quad \text{if } (t, x(t)) \in G_k, \quad k = 1, 2, 3, \\ A_i &= A_i^k \quad \text{if } (t, x(t)) \in G_k, \quad k = 1, 2, 3, \\ \Delta \bar{A}_i(t) &= \begin{cases} 0 & \text{if } (t, x(t)) \in G_1, \\ \Delta A_i(t) & \text{if } (t, x(t)) \in G_2 \cup G_3, \end{cases} \\ \bar{\Lambda}(t) &= \begin{cases} I & \text{if } (t, x(t)) \in G_1 \cup G_2, \\ \Lambda(t) & \text{if } (t, x(t)) \in G_3, \end{cases} \\ \bar{d}(t) &= \begin{cases} 0 & \text{if } (t, x(t)) \in G_1, \\ d(t) & \text{if } (t, x(t)) \in G_2, \\ \varepsilon_u(t) + d(t) & \text{if } (t, x(t)) \in G_3. \end{cases} \end{aligned} \quad (20)$$

Define

$D = \sup_{t \geq 0} \|\varepsilon_u(t) + d(t)\|$ ,  $\vartheta = 1/\inf_{t \geq 0} [\lambda_{\min}(\Lambda(t))]$ . Considering system (19), the control law is designed as follows.

**Controller Rule**  $R_{p,i}$ : IF  $\theta_1(t)$  is  $\mu_{i,1}$ ,  $\theta_2(t)$  is  $\mu_{i,2}$  and ... and  $\theta_p(t)$  is  $\mu_{i,p}$ , THEN

$$\begin{aligned} u(t) &= -\hat{\vartheta} \varphi_{\vartheta}(x(t), v(t)), \\ v(t) &= K_i x(t) - \hat{D} \varphi_D(x(t)), \end{aligned} \quad (21)$$

where  $K_i$  is the control gain matrix to be designed and  $\hat{\vartheta}$  and  $\hat{D}$  are the estimations of  $\vartheta$  and  $D$ , respectively. The continuous functions  $\varphi_{\vartheta}(x(t), v(t))$  and  $\varphi_D(x(t))$  are defined as

$$\begin{aligned}\varphi_{\vartheta}(x(t), v(t)) &= [\varphi_{\vartheta,1}(t), \varphi_{\vartheta,2}(t), \dots, \varphi_{\vartheta,m}(t)]^T, \\ \varphi_{\vartheta,l}(x(t), v(t)) &= \frac{v_l(t)[z_l(t)\widehat{\vartheta}v_l(t)]}{\sqrt{[z_l(t)\widehat{\vartheta}v_l(t)]^2 + \varepsilon_v^2}}, \quad l = 1, 2, \dots, m, \\ \varphi_D(x(t)) &= \frac{B^T P x(t)}{\sqrt{x^T(t) P B B^T P x(t) + \varepsilon_D^2}},\end{aligned}\tag{22}$$

where

$$B = \sum_{i=1}^r h_i(\theta(t)) B_i, \tag{23}$$

$P \in \mathbb{R}^{n \times n}$  is a symmetric positive matrix.  $\varepsilon_v, \varepsilon_D > 0$  are design constants.  $z_l(t)$  is the  $l$ th component of  $z(t) = [x^T(t) P \mathcal{B}]^T$ . The adaptive parameters are updated by

$$\begin{aligned}\dot{\widehat{D}} &= \eta_D x^T(t) P B \varphi_D(x(t)) - \eta_D \sigma_D \widehat{D}, \\ \dot{\widehat{\vartheta}} &= -\eta_{\vartheta} x^T(t) P B v(t) - \eta_{\vartheta} \sigma_{\vartheta} \widehat{\vartheta},\end{aligned}\tag{24}$$

where  $\eta_D, \eta_{\vartheta} > 0$  are the gains of adaptive laws and  $\sigma_D, \sigma_{\vartheta} > 0$  are design constants.

*Remark 2.* Note that it is improper to separately design the control laws for each  $G_k$ ,  $k = 1, 2, 3$ , and a universal controller has to be developed for all the three modes. In practical, the condition of multiple uncertainties suddenly changes or actuator failure abruptly occurs which cannot be determined easily. Moreover, the condition is concerned not only with the system time but also with the system states, which makes this problem more complex. Since the boundaries of  $G_k$  are unknown in practical, the separately design methods cannot be applied and a universal control law which is applicable for all the three modes is necessary. In this paper,  $G_k$ ,  $k = 1, 2, 3$ , are only used for analysis, but are not used in control design.

*Remark 3.* The considered system cannot be controlled by using the controllers of conventional switched systems possessing switching signals those only vary with time. The three modes of the concerned system are distinguished by using the conditions concerned with both the system time and the system states. In this paper, the switchings among the three modes are more intrinsic and are difficult to be dealt with. In fact, the proposed controller can degrade into an asynchronous control law if  $G_k$ ,  $k = 1, 2, 3$  is only concerned with time. For the considered  $G_k$ , the proposed controller can be thought of as a deep asynchronous controller.

**3.2. Stability Analysis.** Define  $\widetilde{\vartheta}(t) = \widehat{\vartheta}(t) - \vartheta(t)$ ,  $\widetilde{D}(t) = \widehat{D}(t) - D(t)$ . Combining (19)–(21) yields

$$\begin{aligned}\dot{x} \in & \sum_{i=1}^r h_i(\theta(t)) \sum_{j=1}^r h_j(\theta(t)) \{F[f_i](x(t)) + F[A_i + B_i K_j] \\ & \cdot (x(t)) + F[\Delta \bar{A}_i(t)](x(t)) + B_i F[\bar{d}(t)] - B_i \widehat{D} \varphi_D \\ & \cdot (x(t)) + B_i F[\bar{\Lambda}(t)](u(t)) - B_i v(t)\}.\end{aligned}\tag{25}$$

**Theorem 1.** Consider the closed-loop fuzzy differential inclusion (25) under Assumptions 1–4. The fuzzy controller is designed as (21) and the adaptive parameters are updated by (24). Given scalars  $\alpha, \lambda > 0$ . For any  $i, j \in Y$  and  $k \in \{1, 2, 3\}$ , if there exist matrices  $P, K$  such that

$$\begin{aligned}\Pi_{i,j}^k &< 0, \quad i = 1, 2, \dots, r, \\ \Pi_{i,j}^k + \Pi_{i,j}^k &< 0, \quad 1 \leq i < j \leq r,\end{aligned}\tag{26}$$

where

$$\Pi_{i,j}^k = \begin{bmatrix} \Pi_{i,j,11}^k & P \\ P & -\frac{1}{\lambda^2} I \end{bmatrix}, \tag{27}$$

$$\begin{aligned}\Pi_{i,j,11}^k &= P(A_i^k + B_i K_j) + (A_i^k + B_i K_j)^T P \\ &+ \frac{1}{\lambda^2} [U_i^k]^T U_i^k + P M_i M_i^T P + N_i^T N_i + \alpha P,\end{aligned}$$

then, for any initial conditions, the Filippov solution of closed-loop fuzzy differential inclusion (25) is bounded and converge to a compact set:

$$\Omega_f = \{x \in \mathbb{R}^n | V(x) \leq \varepsilon_f / \beta\}, \tag{28}$$

where

$$\varepsilon_f = \frac{\sigma_D D^2}{2} + \frac{\sigma_{\vartheta} \vartheta}{2} + D \varepsilon_D + \frac{m \varepsilon_v}{\vartheta}, \tag{29}$$

$$\beta = \min\{\alpha, \eta_D \sigma_D, \eta_{\vartheta}, \sigma_{\vartheta}\}.$$

*Proof.* Select a Lyapunov functional candidate as follows:

$$V(x(t)) = \frac{1}{2} x^T(t) P x(t) + \frac{1}{2 \eta_D} \widetilde{D}^2 + \frac{1}{2 \eta_{\vartheta}} \widetilde{\vartheta}^2, \tag{30}$$

where  $P$  is a positive definite matrix. According to Definition 2, the set-valued Lie derivative of a  $V(x(t))$  can be taken as

$$DV(x(t)) = \bigcap_{\xi \in \partial V(x(t))} \xi^T [\psi(x(t), \widehat{D}, v(t)), \dot{\widehat{D}}, \dot{\widehat{\vartheta}}]^T, \tag{31}$$

where

$$\begin{aligned}
\psi(x(t), \widehat{D}, v(t)) &= \sum_{i=1}^r h_i(\theta(t)) \sum_{j=1}^r h_j(\theta(t)) \{F[f_i](x(t)) \\
&\quad + F[A_i + B_i K_j]x(t) + F[\Delta \bar{A}_i(t)]x(t) \\
&\quad + F[\bar{d}(t)] - B_i \widehat{D} \varphi_D(x(t)) \\
&\quad + B_i F[\bar{\Lambda}(t)]u(t) - B_i v(t)\}.
\end{aligned} \tag{32}$$

By defining  $\pi_{f_i}, \pi_{A_{ij}}, \pi_{\Delta A_i}, \pi_{\Lambda}$ , and  $\pi_d$  such that

$$\begin{aligned}
\pi_{f_i} &\in \begin{cases} f_i^k & \text{if } (t, x(t)) \in G_k, \\ [f_i^{k_1} f_i^{k_2}] & \text{if } (t, x(t)) \in \partial G_{k_1} \cap \partial G_{k_2}, \end{cases} \\
\pi_{A_{ij}} &\in \begin{cases} A_i^k + B_i K_j & \text{if } (t, x(t)) \in G_k, \\ [A_i^{k_1} + B_i K_j, A_i^{k_2} + B_i K_j] & \text{if } (t, x(t)) \in \partial G_{k_1} \cap \partial G_{k_2}, \end{cases} \\
\pi_{\Delta A_i} &\in \begin{cases} 0 & \text{if } (t, x(t)) \in G_1, \\ \Delta A_i & \text{if } (t, x(t)) \in G_2 \cup G_3, \\ [0, \Delta A_i] & \text{if } (t, x(t)) \in \partial G_1 \cap \partial [G_2 \cup G_3], \end{cases} \\
\pi_{\Lambda} &\in \begin{cases} I & \text{if } (t, x(t)) \in G_1 \cup G_2, \\ \Lambda(t) & \text{if } (t, x(t)) \in G_3, \\ [\Lambda(t), I] & \text{if } (t, x(t)) \in \partial [G_1 \cup G_2] \cap \partial G_3, \end{cases} \\
\pi_d &\in \begin{cases} 0 & \text{if } (t, x(t)) \in G_1, \\ d & \text{if } (t, x(t)) \in G_2, \\ \varepsilon_u + d & \text{if } (t, x(t)) \in G_3, \\ [0, d] & \text{if } (t, x(t)) \in \partial G_1 \cap \partial G_2, \\ [d, \varepsilon_u + d] & \text{if } (t, x(t)) \in \partial G_2 \cap \partial G_3, \\ [0, \varepsilon_u + d] & \text{if } (t, x(t)) \in \partial G_1 \cap \partial G_3, \end{cases}
\end{aligned} \tag{33}$$

we can rewrite (30) as

$$DV(x(t)) = D_1 V(x(t)) + D_2 V(x(t)), \tag{34}$$

where

$$D_1 V(x(t)) = \sum_{i=1}^r h_i(\theta(t)) \sum_{j=1}^r h_j(\theta(t)) \left\{ x^T(t) P \pi_{f_i} + x^T(t) P \pi_{A_{ij}} x(t) + x^T(t) P \pi_{\Delta A_i} x(t) \right\}, \tag{35}$$

$$D_2 V(x(t)) = \sum_{i=1}^r h_i(\theta(t)) \left\{ x^T(t) P B_i \pi_d - x^T(t) P B_i \widehat{D} \varphi_D(x(t)) + x^T(t) P B_i \pi_{\Lambda} u(t) - x^T(t) P B_i v(t) \right\} + \frac{1}{\eta_D} \dot{\widehat{D}} + \frac{1}{\eta_{\theta}} \dot{\widehat{\theta}}. \tag{36}$$

Firstly, the analysis of  $\mathcal{D}_1 V(x(t))$  are given. Since  $P$  is nonsingular, it can be proved that  $P \pi_{\Delta A_i} \in [0, P \Delta A_i]$  holds for all  $(t, x(t))$ . Hence, based on Assumption 2, it can be known that, for any  $(t, x(t))$ ,  $x^T(t) P \pi_{\Delta A_i} x(t) \in [0, x^T(t) P \Delta A_i x(t)]$  and

$$x^T(t) P \pi_{\Delta A_i} x(t) \leq \frac{1}{2} x^T(t) P M_i M_i^T P x(t) + \frac{1}{2} x^T(t) N_i^T N_i x(t). \tag{37}$$

For  $(t, x(t)) \in G_k, k = 1, 2, 3$ , it follows from Assumption 1 that

$$\begin{aligned} \pi_{f_i}^T \pi_{f_i} &\leq x^T(t) [U_i^k]^T U_i^k x(t), \\ P\pi_{A_{ij}} &= \frac{1}{2} \left[ P(A_i^k + B_i K_j) + (A_i^k + B_i K_j)^T P \right]. \end{aligned} \quad (38)$$

By combining (35), (37), and (38), we know that, for  $(t, x(t)) \in G_k$ ,  $k = 1, 2, 3$ , the following inequality holds:

$$\begin{aligned} \Phi_{i,j}^k &= \begin{bmatrix} \Phi_{i,j,11}^k & P \\ P & -\frac{1}{\lambda^2} \end{bmatrix}, \\ \Phi_{i,j,11}^k &= P(A_i^k + B_i K_j) + (A_i^k + B_i K_j)^T P + \frac{1}{\lambda^2} [U_i^k]^T U_i^k \\ &\quad + PM_i M_i^T P + N_i^T N_i. \end{aligned} \quad (40)$$

By using (26), it can be checked that, for  $(t, x(t)) \in G_k$ ,  $k = 1, 2, 3$ ,

$$D_1 V(x(t)) \leq -\frac{\alpha}{2} x^T(t) P x(t). \quad (41)$$

where

On the contrary, for  $(t, x(t)) \in \partial G_{k_1} \cap \partial G_{k_2}$ ,  $k_1, k_2 \in \{1, 2, 3\}$ , it can be proved that

$$\begin{aligned} x^T(t) P \pi_{A_{ij}} x(t) &\in [x^T(t) P (A_i^{k_1} + B_i K_j) x(t), x^T(t) P (A_i^{k_2} + B_i K_j) x(t)], \\ \pi_{f_i}^T \pi_{f_i} &\in [f_i^{k_1}]^T f_i^{k_1}, [f_i^{k_2}]^T f_i^{k_2} \in [x^T(t) [U_i^{k_1}]^T U_i^{k_1} x(t), x^T(t) [U_i^{k_2}]^T U_i^{k_2} x(t)]. \end{aligned} \quad (42)$$

Accordingly, the following inequalities can be obtained:

$$\begin{aligned} x^T(t) P \pi_{A_{ij}} x(t) &\leq \max_{k \in \{1,2,3\}} x^T(t) P (A_i^k + B_i K_j) x(t), \\ \pi_{f_i}^T \pi_{f_i} &\leq \max_{k \in \{1,2,3\}} x^T(t) [U_i^k]^T U_i^k x(t). \end{aligned} \quad (43)$$

By combining (35), (37), and (43), we know that, for  $(t, x(t)) \in \partial G_{k_1} \cap \partial G_{k_2}$ ,  $k_1, k_2 \in \{1, 2, 3\}$ ,

$$D_1 V(x(t)) \leq \frac{1}{2} \sum_{i=1}^r h_i(\theta(t)) \sum_{j=1}^r h_j(\theta(t)) \begin{bmatrix} x(t) \\ \pi_{f_i} \end{bmatrix}^T \Psi_{i,j}^k \begin{bmatrix} x(t) \\ \pi_{f_i} \end{bmatrix}, \quad (44)$$

where

$$\begin{aligned} \Psi_{i,j}^k &= \begin{bmatrix} \Psi_{i,j,11}^k & P \\ P & -\frac{1}{\lambda^2} \end{bmatrix}, \\ \Psi_{i,j,11}^k &= \max_{k \in \{1,2,3\}} \left\{ P(A_i^k + B_i K_j) + (A_i^k + B_i K_j)^T P \right. \\ &\quad \left. + \frac{1}{\lambda^2} [U_i^k]^T U_i^k \right\} + PM_i M_i^T P + N_i^T N_i. \end{aligned} \quad (45)$$

Since (26) holds for any  $k \in \{1, 2, 3\}$ , we know that, for  $(t, x(t)) \in \partial G_{k_1} \cap \partial G_{k_2}$ ,  $k_1, k_2 \in \{1, 2, 3\}$ ,  $\mathcal{D}_1 V(x(t)) \leq -\alpha/2 x^T(t) P x(t)$  also holds. Then, by considering (41), it can be concluded that, for any  $(t, x(t))$ ,

$$D_1 V(x(t)) \leq -\frac{\alpha}{2} x^T(t) P x(t). \quad (46)$$

Next texts provide the analysis of  $\mathcal{D}_2 V(x(t))$ . From (33), we know that, for  $(t, x(t))$ ,

$$\begin{aligned} \pi_d &\in [0, \varepsilon_u + d], \\ \pi_{\Lambda_i} &\in [\Lambda_i(t), 1], \end{aligned} \quad (47)$$

where  $\pi_{\Lambda_i}$  and  $\Lambda_i(t)$  represent the  $i$ th component of  $\pi_\Lambda$  and  $\Lambda(t)$  on the diagonal line, respectively. Hence, it can be obtained that, for any  $(t, x(t))$ ,

$$\begin{aligned} \sum_{i=1}^r h_i(\theta(t)) x^T(t) P B_i \pi_d &\in [0, x^T(t) P B (\varepsilon_u + d)], \\ \sum_{i=1}^r h_i(\theta(t)) x^T(t) P B_i \pi_{\Lambda_i} u(t) &\in [x^T(t) P B \Lambda(t) u(t), x^T(t) P B u(t)]. \end{aligned} \quad (48)$$

From Lemma 2, it is easy to know that

$$\begin{aligned}
& \sum_{i=1}^r h_i(\theta(t)) \{x^T(t)PB_i\pi_d - x^T(t)PB_i\widehat{D}\varphi_D(x(t))\} \\
& \leq D\|x^T(t)PB\| - x^T(t)PB\widehat{D}\varphi_D(x(t)) \\
& \leq D \frac{x^T(t)PB[x^T(t)PB]^T}{\sqrt{x^T(t)PB[x^T(t)PB]^T + \varepsilon_D^2}} - x^T(t)PB\widehat{D}\varphi_D(x(t)) + D\varepsilon_D \\
& \leq -\widehat{D}x^T(t)PB\varphi_D(x(t)) + D\varepsilon_D.
\end{aligned} \tag{49}$$

Meanwhile, simple computation shows that

$$\begin{aligned}
& \sum_{i=1}^r h_i(\theta(t)) \{x^T(t)PB_i\pi_{\Lambda}u(t) - x^T(t)PB_iv(t)\} \\
& = -\widehat{\vartheta} \sum_{l=1}^m z_l\pi_{\Lambda,l}\varphi_{\vartheta,l}(x(t), v(t)) - x^T(t)PBv(t).
\end{aligned} \tag{50}$$

Since  $0 < 1/\vartheta \leq \pi_{\Lambda,i} \in [\Lambda_i(t), 1]$ , we can get that

$$\begin{aligned}
& -\widehat{\vartheta}z_l\pi_{\Lambda,l}\varphi_{\vartheta,l}(x(t), v(t)) \\
& = \frac{z_l\pi_{\Lambda,l}\widehat{\vartheta}v_l(t)[z_l(t)\widehat{\vartheta}v_l(t)]}{\sqrt{[z_l(t)\widehat{\vartheta}v_l(t)]^2 + \varepsilon_v^2}} \\
& \leq -\frac{1}{\widehat{\vartheta}} \frac{[z_l(t)\widehat{\vartheta}v_l(t)]^2}{\sqrt{[z_l(t)\widehat{\vartheta}v_l(t)]^2 + \varepsilon_v^2}}.
\end{aligned} \tag{51}$$

Hence, it follows from Lemma 2 that

$$\begin{aligned}
& -\widehat{\vartheta} \sum_{l=1}^m z_l\pi_{\Lambda,l}\varphi_{\vartheta,l}(x(t), v(t)) - x^T(t)PBv(t) \\
& \leq \sum_{l=1}^m \left[ -\frac{1}{\widehat{\vartheta}} \frac{[z_l(t)\widehat{\vartheta}v_l(t)]^2}{\sqrt{[z_l(t)\widehat{\vartheta}v_l(t)]^2 + \varepsilon_v^2}} - z_l(t)v_l(t) \right] \\
& \leq \sum_{l=1}^m \left[ \frac{1}{\widehat{\vartheta}} \left( -\|z_l(t)\widehat{\vartheta}v_l(t)\| + \varepsilon_v \right) - z_l(t)v_l(t) \right] \\
& \leq \sum_{l=1}^m \left[ \frac{\widehat{\vartheta}}{\widehat{\vartheta}^2} z_l(t)v_l(t) - z_l(t)v_l(t) + \frac{\varepsilon_v}{\widehat{\vartheta}} \right] \\
& \leq \frac{\widehat{\vartheta}}{\widehat{\vartheta}^2} x^T(t)PBv(t) + \frac{m\varepsilon_v}{\widehat{\vartheta}}.
\end{aligned} \tag{52}$$

By combining (36) and (49)–(52), it can be obtained that

$$\begin{aligned}
D_2V(x(t)) & \leq -\widehat{D}x^T(t)PB\varphi_D(x(t)) + \frac{\widehat{\vartheta}}{\vartheta}x^T(t)PBv(t) \\
& \quad + \frac{1}{\eta_D}\widehat{D}\dot{\widehat{D}} + \frac{1}{\eta_D\vartheta}\widehat{\vartheta}\dot{\widehat{\vartheta}} + D\varepsilon_D + \frac{m\varepsilon_v}{\vartheta},
\end{aligned} \tag{53}$$

where  $m$  is the dimension of the system input signal  $u(t)$ . Substituting (24) into (53) yields

$$D_2V(x(t)) \leq -\sigma_D\widehat{D}\widehat{D} - \frac{\sigma_D\widehat{\vartheta}\widehat{\vartheta}}{\vartheta} + D\varepsilon_D + \frac{m\varepsilon_v}{\vartheta}. \tag{54}$$

By using the following inequalities,

$$\begin{aligned}
2\widehat{D}\widehat{D} & \leq -\widehat{D}^2 + D^2, \\
\widehat{\vartheta}\widehat{\vartheta} & \leq -\widehat{\vartheta}^2 + \vartheta^2.
\end{aligned} \tag{55}$$

We know that

$$D_2V(x(t)) \leq -\frac{\sigma_D}{2}\widehat{D}^2 - \frac{\sigma_D\widehat{\vartheta}^2}{2\vartheta} + \varepsilon_f. \tag{56}$$

By combining (34), (46), and (56), we know that

$$DV(x(t)) \leq -\frac{\alpha}{2}x^T(t)Px(t) - \frac{\sigma_D}{2}\widehat{D}^2 - \frac{\sigma_D\widehat{\vartheta}^2}{2\vartheta} + \varepsilon_f, \tag{57}$$

which means

$$DV(x(t)) \leq -\beta V(x(t)) + \varepsilon_f, \tag{58}$$

where  $\beta$  and  $\varepsilon_f$  are defined in (29). According to Lemma 1, it can be proved that the Filippov solution of closed-loop fuzzy differential inclusion (25) is bounded and converge to  $\Omega_f$ . The proof is complete.  $\square$

*Remark 4.* It should be noted that the final compact set  $\Omega_f$  can be an arbitrarily small neighborhood of the origin by adjusting the control gains and adaptive parameters. Moreover, in most of the adaptive control results, the ultimate boundness of the closed-loop control system is finally ensured. However, for the differential inclusion systems, the criteria of ultimately boundness have never been provided. Hence, it can be concluded that Lemma 1 lays the mathematical fundamentals for the adaptive control of differential inclusion systems.

Next, we will provide the computation method of control gain  $K$  and matrix  $P$  which is necessary in adaptive laws.

**Theorem 2.** Consider the closed-loop fuzzy differential inclusion (25) under Assumptions 1–4. The fuzzy controller is designed as (20), and the adaptive parameters are updated by (24). Given scalars  $\alpha, \lambda > 0$ . For any  $i, j \in Y$  and  $k \in \{1, 2, 3\}$ , if there exist matrices  $Q, R_j$  such that

$$\begin{aligned}
\Gamma_{i,j}^k & < 0, \quad i = 1, 2, \dots, r, \\
\Gamma_{i,j}^k + \Gamma_{i,j}^k & < 0, \quad 1 \leq i < j \leq r,
\end{aligned} \tag{59}$$

where

$$\Gamma_{i,j}^k = \begin{bmatrix} \Gamma_{i,j,11}^k & Q[U_i^k]^T & QN_i^T & I \\ * & -\lambda^2 I & 0 & 0 \\ * & * & -I & 0 \\ * & * & * & \frac{1}{\lambda^2} I \end{bmatrix}, \quad (60)$$

$$\Gamma_{i,j,11}^k = A_i^k Q + Q A_i^k + B_i R_j + R_j^T B_i^T + \alpha Q + M_i M_i^T,$$

then, for any initial conditions, the Filippov solution of closed-loop fuzzy differential inclusion (25) is bounded and converges to the compact set  $\Omega_f$ . Moreover, if condition (59) is feasible, it can be obtained that

$$\begin{aligned} K_j &= R_j Q^{-1}, \\ P &= Q^{-1}. \end{aligned} \quad (61)$$

*Proof.* Define  $Q = P^{-1}$  and  $R_j = K_j P^{-1}$ . By performing a congruence transformation to with  $\Pi_{i,j}^k$ , we can get the following matrix:

$$\widehat{\Pi}_{i,j}^k = \begin{bmatrix} \widehat{\Pi}_{i,j,11}^k & I \\ I & -\frac{1}{\lambda^2} I \end{bmatrix}, \quad (62)$$

where

$$\begin{aligned} \widehat{\Pi}_{i,j,11}^k &= A_i^k Q + Q A_i^k + B_i R_j + R_j^T B_i^T + \alpha Q + \frac{1}{\lambda^2} Q [U_i^k]^T U_i^k Q \\ &+ M_i M_i^T + Q N_i^T N_i Q. \end{aligned} \quad (63)$$

By performing a Schur complement transformation to (61),  $\Gamma_{i,j}^k$  can be obtained. Since

$$\Gamma_{i,j}^k < 0 \Leftrightarrow \widehat{\Pi}_{i,j}^k < 0 \Leftrightarrow \Pi_{i,j}^k < 0, \quad (64)$$

we know that (26) is satisfied by using (58). According to Theorem 1, the boundedness and convergence of the Filippov solution can be guaranteed. The proof is complete.  $\square$

*Remark 5.* System who involves state-based switching usually has more complex dynamic behaviors which motivates various useful applications, while the construction of rigorous stability for these systems is challenging. Different to classic analysis methods in switching systems such as average dwell time (ADT), differential-inclusion-based approaches provide feasible alternative solutions to the stability analysis of stated-based switching systems. Instead of

analyzing the value of the vector field at individual points, differential-inclusion-based methods focus on the behavior of vector field at the neighborhood of each point. This idea is at the core of constructing continuous Filippov solution which is a general solution to the original differential equations, where the discontinuities is covered by set-valued mapping which is a convex combination of vector field around them. As a result, rigorous stability result can be established for discontinuous vector field as long as the Lebesgue measure of the discontinuities is zero. From the illustration above, it is clear that, for state-based switching systems, the superiority of utilizing differential inclusion is significant.

#### 4. Simulation Study

In this section, we will present a numerical example with two fuzzy subsystems to demonstrate the effectiveness of the proposed control method. The switching regions  $G_i$  of system (1) are defined as

$$\begin{aligned} G_1 &= \{x \in \mathbb{R}^2 \mid \|x\| \geq n1\}, \\ G_2 &= \{x \in \mathbb{R}^2 \mid 0.5t \leq n\|x\|q < h1\}, \\ G_3 &= \{x \in \mathbb{R}^2 \mid \|x\| < n0.5\}. \end{aligned} \quad (65)$$

The system-related matrices are given as follows:

$$\begin{aligned} A_1^1 &= \begin{bmatrix} -2.7 & 1.5 \\ 0.5 & 1.2 \end{bmatrix}, \\ A_2^1 &= \begin{bmatrix} -2.1 & 0.6 \\ 0.2 & 1.8 \end{bmatrix}, \\ A_1^2 &= \begin{bmatrix} -1.9 & 2.3 \\ 1.5 & 2.8 \end{bmatrix}, \\ A_2^2 &= \begin{bmatrix} -3.6 & 2.7 \\ 0.1 & 1.5 \end{bmatrix}, \\ A_1^3 &= \begin{bmatrix} -1.2 & 0.7 \\ 2.1 & 0.5 \end{bmatrix}, \\ A_2^3 &= \begin{bmatrix} -2.8 & 1.7 \\ 3.5 & 1.9 \end{bmatrix}, \\ B_1 &= \begin{bmatrix} -1.9 \\ 3.6 \end{bmatrix}, \\ B_2 &= \begin{bmatrix} -1.5 \\ 2.3 \end{bmatrix}. \end{aligned} \quad (66)$$

The unknown nonlinear functions are set as

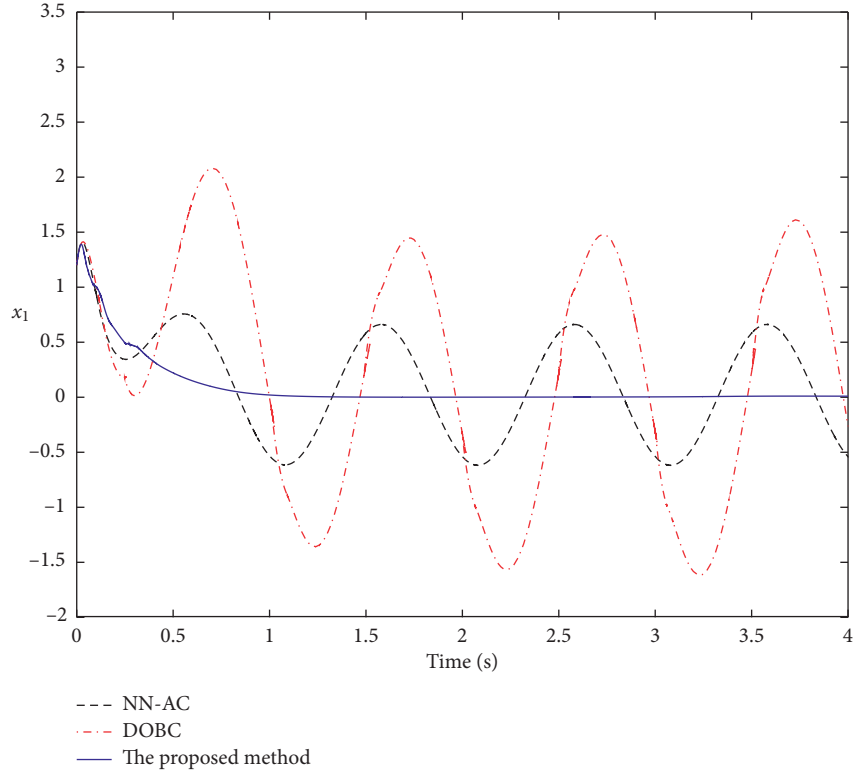


FIGURE 1: The trajectories of the system state  $x_1(t)$  under NN-AC, DOBC, and proposed method.

$$\begin{aligned}
 f_1^1 &= \begin{bmatrix} 0.016 \cos(2x_1x_2) \\ 0.022 \sin(x_1 + x_2) \end{bmatrix}, \\
 f_2^1 &= \begin{bmatrix} 0.018 \sin(x_1^2) \\ 0.02 \cos(x_2^2) \end{bmatrix}, \\
 f_1^2 &= \begin{bmatrix} 0.06 \sin(x_1)\cos(x_2) \\ 0.025 \sin(\|x_1\|) \end{bmatrix}, \\
 f_2^2 &= \begin{bmatrix} 0.028 \cos(x_1)\cos(x_2) \\ 0.025 \cos(\|x_2\|) \end{bmatrix}, \\
 f_1^3 &= \begin{bmatrix} 0.024 \sin(x_1)\cos(x_2) \\ 0.025 \cos(\|x_1\| + \|x_2\|) \end{bmatrix}, \\
 f_2^3 &= \begin{bmatrix} 0.01 \sin(x_1x_2) \\ 0.05 \cos(\|x_1\|x_2) \end{bmatrix}.
 \end{aligned} \tag{67}$$

$$\begin{aligned}
 M_1 &= \begin{bmatrix} 0.012 & 0 \\ 0 & 0.08 \end{bmatrix}, \\
 M_2 &= \begin{bmatrix} 0.015 & 0 \\ 0 & 0.06 \end{bmatrix}, \\
 N_1 &= \begin{bmatrix} 0.069 & 0 \\ 0 & 0.087 \end{bmatrix}, \\
 N_2 &= \begin{bmatrix} 0.055 & 0 \\ 0 & 0.072 \end{bmatrix}, \\
 U_1^1 = U_1^2 = U_1^3 &= \begin{bmatrix} 0.25 & 0 \\ 0 & 0.25 \end{bmatrix}, \\
 U_2^1 = U_2^2 = U_2^3 &= \begin{bmatrix} 0.3 & 0 \\ 0 & 0.3 \end{bmatrix}.
 \end{aligned} \tag{68}$$

By letting  $\alpha = 2$  and  $\lambda = 1$ , we can solve (59) and obtain that

$$\begin{aligned}
 P &= \begin{bmatrix} 4.0590 & 6.9122 \\ 6.9122 & 14.0675 \end{bmatrix}, \\
 K_1 &= \begin{bmatrix} -5.6326 & -10.8441 \end{bmatrix}, \\
 K_2 &= \begin{bmatrix} -9.0548 & -17.6541 \end{bmatrix}.
 \end{aligned} \tag{69}$$

To compute the control gains, we take



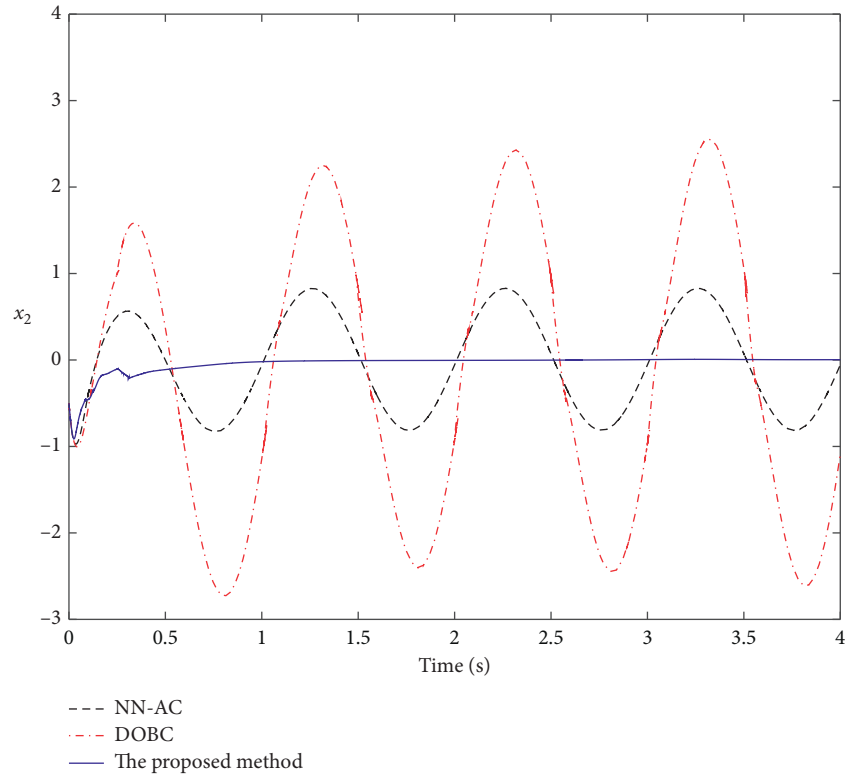


FIGURE 2: The trajectories of the system state  $x_2(t)$  under NN-AC, DOBC, and proposed method.

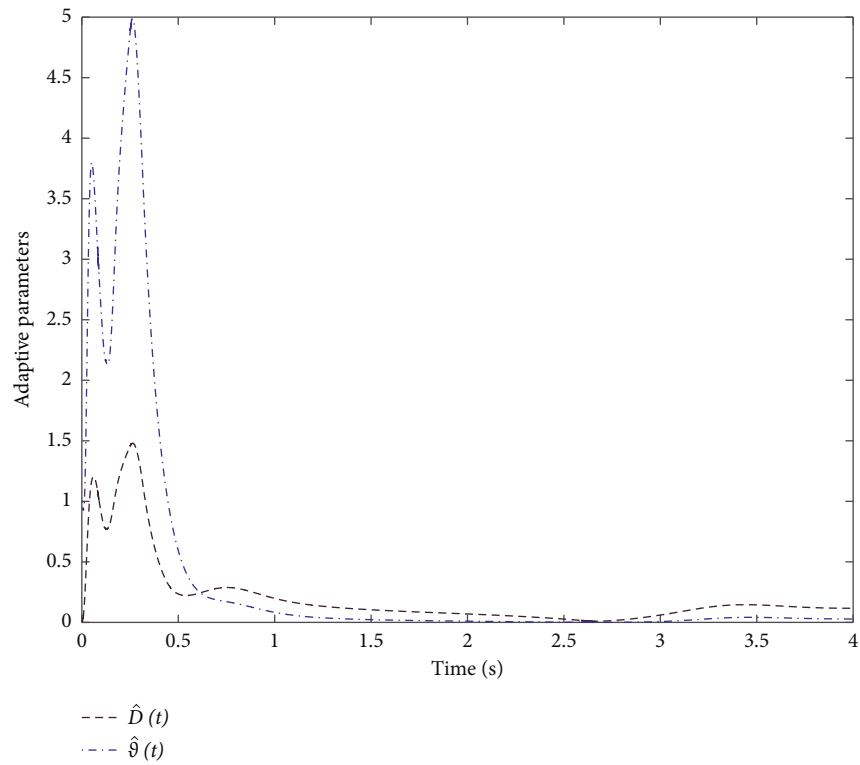
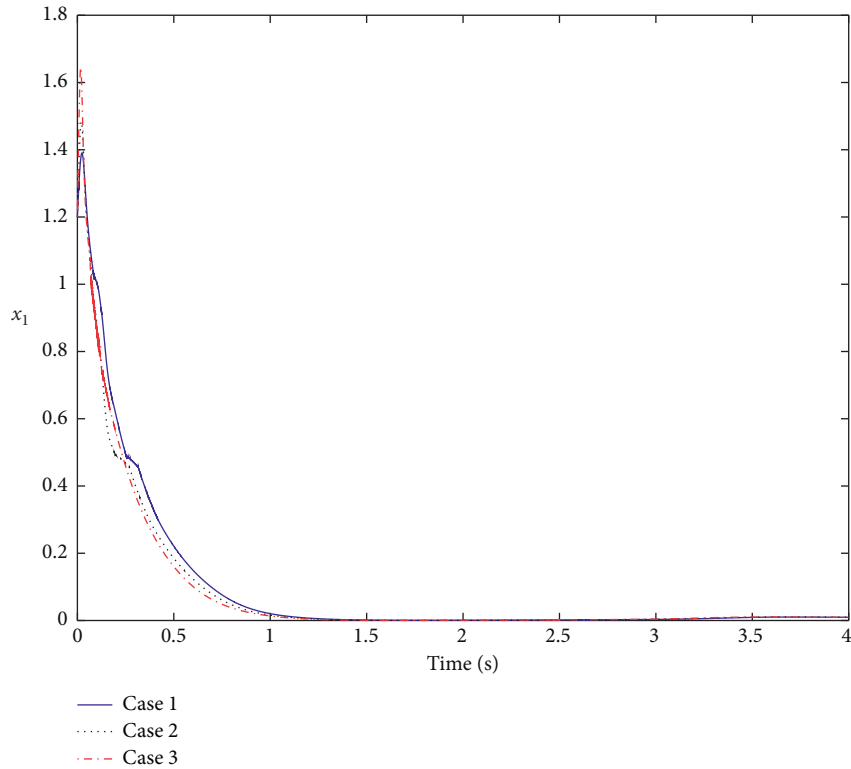


FIGURE 3: The trajectories of the adaptive parameters under the proposed method.

TABLE 2: The parameters of the three simulation cases.

|        | Switching regions  | Fault parameters                            | Disturbance parameters                 |
|--------|--|---|--|
| Case 1 | $G_1 = \{x(t) \in \mathbb{R}^2 \mid \ x\  \geq 1\}$<br>$G_2 = \{x(t) \in \mathbb{R}^2 \mid 0.5 \leq \ x\  \leq 1\}$<br>$G_3 = \{x(t) \in \mathbb{R}^2 \mid \ x\  \leq 0.5\}$     | $\Lambda(t) = 0.7 + 0.1 \sin(t/2 + t\pi/3)$ | $d(t) = 20 \sin(2\pi t - \pi/6) + 0.5$ |
| Case 2 | $G_1 = \{x(t) \in \mathbb{R}^2 \mid \ x\  \geq 1.2\}$<br>$G_2 = \{x(t) \in \mathbb{R}^2 \mid 0.5 \leq \ x\  \leq 1.2\}$<br>$G_3 = \{x(t) \in \mathbb{R}^2 \mid \ x\  \leq 0.5\}$ | $\Lambda(t) = 0.5 + 0.1 \sin(t/2 + t\pi/3)$ | $d(t) = 30 \sin(2\pi t - \pi/6) + 0.5$ |
| Case 3 | $G_1 = \{x(t) \in \mathbb{R}^2 \mid \ x\  \geq 1.2\}$<br>$G_2 = \{x(t) \in \mathbb{R}^2 \mid 0.8 \leq \ x\  \leq 1.2\}$<br>$G_3 = \{x(t) \in \mathbb{R}^2 \mid \ x\  \leq 0.8\}$ | $\Lambda(t) = 0.3 + 0.1 \sin(t/2 + t\pi/3)$ | $d(t) = 50 \sin(2\pi t - \pi/6) + 0.5$ |

FIGURE 4: The trajectories of the system state  $x_1(t)$  of the proposed method under three cases.

In the simulation, the initial values of the system states and adaptive parameters are set as  $x_1(0) = 1.2$ ,  $x_2(0) = -0.5$ ,  $\hat{D}(0) = 0$ , and  $\hat{\vartheta}(0) = 1$ . The corresponding adaptive gains and constants are selected as  $\eta_D = \eta_{\vartheta} = 5$ ,  $\sigma_D = \sigma_{\vartheta} = 2$ , and  $\varepsilon_D = \varepsilon_{\nu} = 0.001$ . The disturbance is taken as

$$d(t) = 20 \sin\left(2\pi t - \frac{\pi}{6}\right) + 0.5. \quad (70)$$

The actuator-related settings are

$$\Lambda(t) = 0.7 + 0.1 \sin\left(\frac{t}{2} + \frac{\pi}{3}\right); \varepsilon_u = 0.001. \quad (71)$$

Moreover, to reveal the advantages of the proposed method, the advanced neural network based adaptive control method (NN-AC) and the disturbance observer based control method (DOBC) have been employed in the simulation experiments. The parameters of the NN-AC method are set as  $\eta_{\Theta} = 5$  and  $\sigma_{\Theta} = 2$ . The disturbance observation gain of the DOBC method is selected as  $L = [1.2, 0; 0, 1.2]$ . The control gains of the NN-AC and DOBC methods are set as the same as the proposed method.

The simulation results are provided in Figures 1–3. It is obvious that the proposed method can force both  $x_1(t)$  and  $x_2(t)$  to converge towards zero under switching between  $G_i$

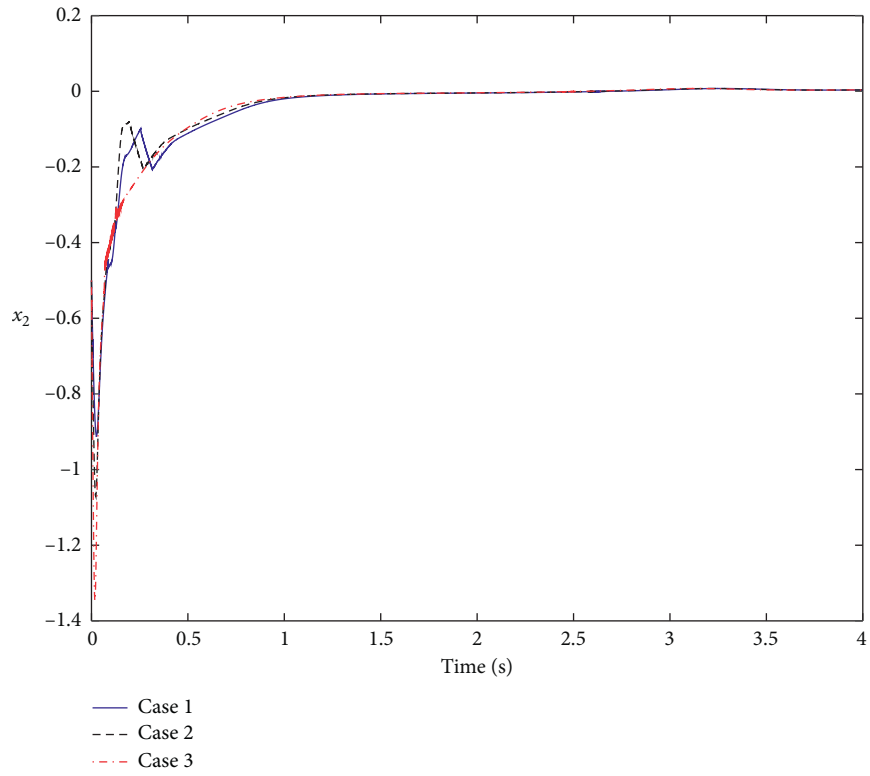


FIGURE 5: The trajectories of the system state  $x_2(t)$  of the proposed method under three cases.

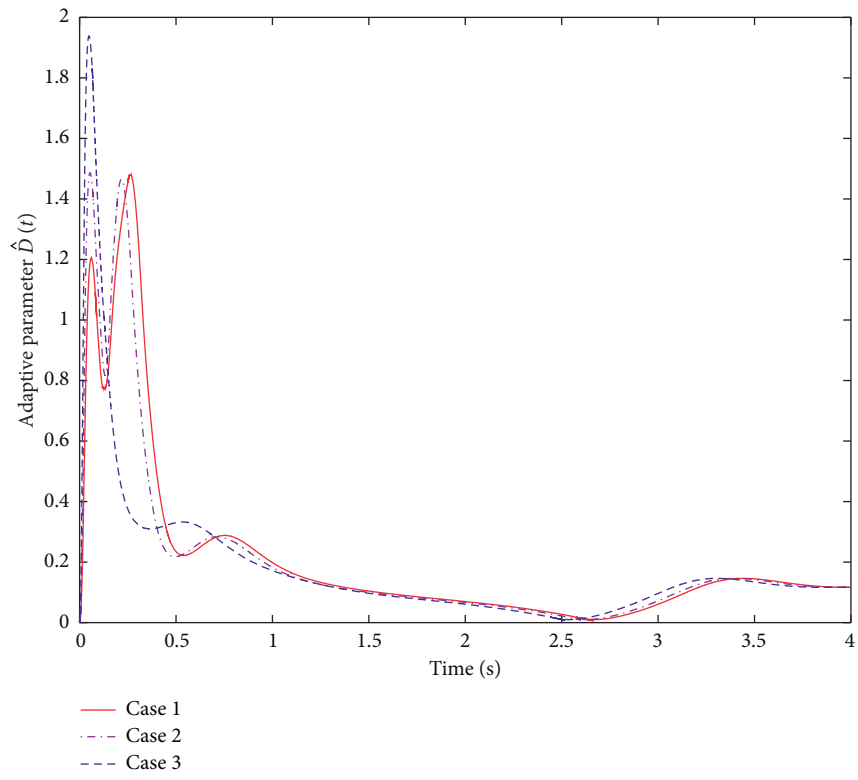


FIGURE 6: The trajectories of the system state  $\hat{D}(t)$  of the proposed method under three cases.

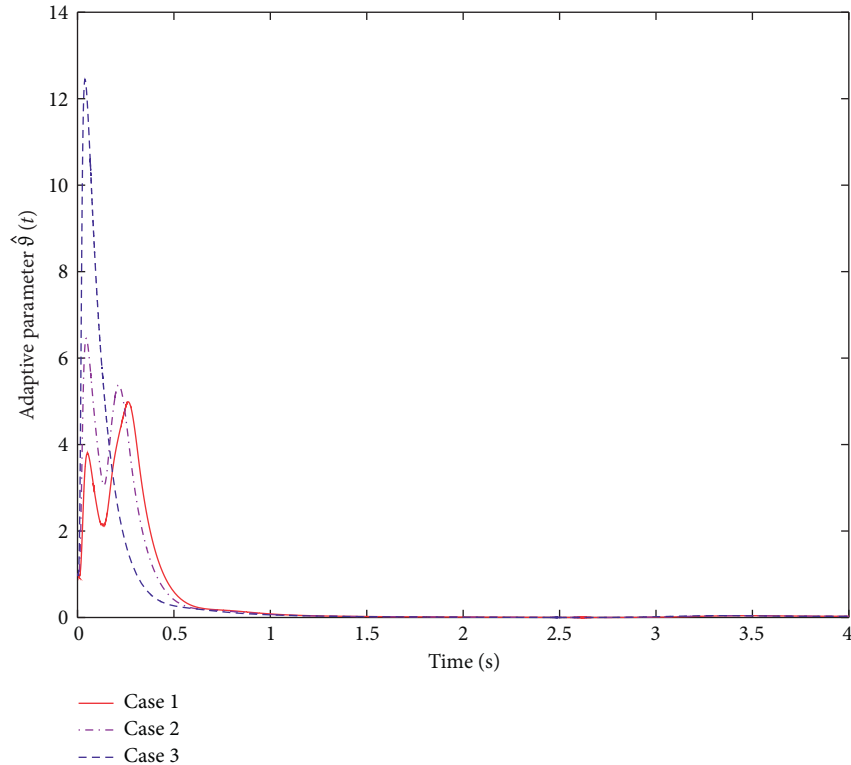


FIGURE 7: The trajectories of the system state  $\hat{\vartheta}(t)$  of the proposed method under three cases.

and finally stay inside a small region near the equilibrium point. Differently, the NN-AC and the DOBC method will cause violent shock, demonstrating that these two methods may possess worse adaptability for the state-dependent switching regions compared with the proposed method. Moreover, it can be found that, by using the proposed method, the adaptive parameters  $\hat{D}(t)$  and  $\hat{\vartheta}(t)$  also converge to a stable value after the initial transient. It can be concluded that, using the proposed method, the closed-loop stability can be guaranteed even under the worst situations (where disturbances and actuator faults both occur). The advantages of the proposed method can be revealed therefore.

Furthermore, to show the robustness of the proposed method, three cases are considered. The parameters of the three cases are given by Table 2.

Under the three cases, the simulation results using the proposed method are given in Figures 4–7. It can be found that although the switching regions, the fault parameters, and the disturbance parameters have changed, the proposed method can still achieve desired control performance.

## 5. Conclusions

A novel robust adaptive controller is given in this paper for solving one of the motivating problems in nonlinear fuzzy systems, that is, to appropriately describe the behavior of the system and to guarantee the stability of the system under discontinuous multiple uncertainties and state-based switching. The proposed differential-inclusion-based

method provides a constructive procedure for the controller design and analysis of a class of heterogeneous T-S fuzzy nonlinear systems with suddenly changing structural parameters and abrupt actuator failures where switching of system dynamics is related with both time and system states. The stability of the resulting closed-loop differential inclusion system is rigorously discussed by virtue of introducing a new fundamental stability lemma for adaptive discontinuous systems, and our results are validated by carefully designed simulations. It should be noted that our control scheme can be easily extended to other T-S fuzzy nonlinear systems with discontinuities and state-based switching, which may provide useful insights for further future research.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant no. 11772256 and sponsored by Innovation Foundation for Doctor Dissertation of Northwestern Polytechnical University (G2017KY0412).

## References

- [1] J. Cheng, J. H. Park, L. Zhang, and Y. Zhu, "An asynchronous operation approach to event-triggered control for fuzzy markovian jump systems with general switching policies," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 1, pp. 6–18, 2018.
- [2] J. Lian, S. Li, and J. Liu, "T-S fuzzy control of positive markov jump nonlinear systems," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 4, pp. 2374–2383, 2018.
- [3] H. Shen, F. Li, Z.-G. Wu, J. H. Park, and V. Sreeram, "Fuzzy-model-based nonfragile control for nonlinear singularly perturbed systems with semi-markov jump parameters," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 6, pp. 3428–3439, 2018.
- [4] X.-H. Chang and G.-H. Yang, "Nonfragile  $H_{\infty}$  filtering of continuous-time fuzzy systems," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1528–1538, 2011.
- [5] H. D. Tuan, P. Apkarian, T. Narikiyo, and Y. Yamamoto, "Parameterized linear matrix inequality techniques in fuzzy control system design," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 2, pp. 324–332, 2001.
- [6] H.-K. Lam, "Output regulation of polynomial fuzzy model-based control systems," *Polynomial Fuzzy Model-Based Control Systems*, vol. 64, no. 2, pp. 137–173, 2016.
- [7] Y. Wang, H. Shen, H. R. Karimi, and D. Duan, "Dissipativity-based fuzzy integral sliding mode control of continuous-time T-S fuzzy systems," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 3, pp. 1164–1176, 2018.
- [8] H. Li, J. Yu, C. Hilton, and H. Liu, "Adaptive sliding-mode control for nonlinear active suspension vehicle systems using T-S fuzzy approach," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 8, pp. 3328–3338, 2013.
- [9] B. Jiang, Z. Gao, P. Shi, and Y. Xu, "Adaptive fault-tolerant tracking control of near-space vehicle using takagi-sugeno fuzzy models," *IEEE Transactions on Fuzzy Systems*, vol. 18, no. 5, pp. 1000–1007, 2010.
- [10] J. Tao, R. Lu, H. Su, P. Shi, and Z.-G. Wu, "Asynchronous filtering of nonlinear markov jump systems with randomly occurred quantization via T-S fuzzy models," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 4, pp. 1866–1877, 2017.
- [11] R. Sakthivel, T. Saravanakumar, B. Kaviarasan, and Y. Lim, "Finite-time dissipative based fault-tolerant control of takagi-sugeno fuzzy systems in a network environment," *Journal of the Franklin Institute*, vol. 354, no. 8, pp. 3430–3454, 2017.
- [12] X. Su, Y. Wen, Y.-D. Song, and T. Hayat, "Dissipativity-based fuzzy control of nonlinear systems via an event-triggered mechanism," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 6, pp. 1208–1217, 2019.
- [13] J. Cheng, J. H. Park, and H. Wang, "Event-triggered  $H_{\infty}$  control for T-S fuzzy nonlinear systems and its application to truck-trailer system," *ISA Transactions*, vol. 65, pp. 62–71, 2016.
- [14] P. Shi, X. Su, and F. Li, "Dissipativity-based filtering for fuzzy switched systems with stochastic perturbation," *IEEE Transactions on Automatic Control*, vol. 61, no. 6, pp. 1694–1699, 2016.
- [15] M. Chadli, A. Abdo, and S. X. Ding, "Fault detection filter design for discrete-time Takagi-Sugeno fuzzy system," *Automatica*, vol. 49, no. 7, pp. 1996–2005, 2013.
- [16] Y. Zhao, H. Gao, J. Lam, and B. Du, "Stability and stabilization of delayed T-S fuzzy systems: a delay partitioning approach," *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 4, pp. 750–762, 2009.
- [17] X. Yang, L. Wu, H.-K. Lam, and X. Su, "Stability and stabilization of discrete-time T-S fuzzy systems with stochastic perturbation and time-varying delay," *IEEE Transactions on Fuzzy Systems*, vol. 22, no. 1, pp. 124–138, 2014.
- [18] X.-H. Chang, Q. Liu, Y.-M. Wang, and J. Xiong, "Fuzzy peak-to-peak filtering for networked nonlinear systems with multipath data packet dropouts," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 3, pp. 436–446, 2019.
- [19] B.-Z. Guo, "Active disturbance rejection control: from ODEs to PDEs\*\*This work was carried out with the support of the national natural science foundation of China and the national research foundation of South Africa," *IFAC-PapersOnLine*, vol. 49, no. 8, pp. 278–283, 2016.
- [20] T. Jiang, C. Huang, and L. Guo, "Control of uncertain nonlinear systems based on observers and estimators," *Automatica*, vol. 59, pp. 35–47, 2015.
- [21] S. Li, J. Yang, W.-H. Chen, and X. Chen, *Disturbance Observer-Based Control: Methods and Applications*, CRC Press, Boca Raton, FL, USA, 2014.
- [22] W.-H. Chen, J. Yang, L. Guo, and S. Li, "Disturbance-observer-based control and related methods-an overview," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 2, pp. 1083–1095, 2016.
- [23] L. Guo and W.-H. Chen, "Disturbance attenuation and rejection for systems with nonlinearity via DOBC approach," *International Journal of Robust and Nonlinear Control*, vol. 15, no. 3, pp. 109–125, 2005.
- [24] W.-H. Chen, "Disturbance observer based control for nonlinear systems," *IEEE/ASME Transactions on Mechatronics*, vol. 9, no. 4, pp. 706–710, 2004.
- [25] L. Guo and S. Cao, "Anti-disturbance control theory for systems with multiple disturbances: a survey," *ISA Transactions*, vol. 53, no. 4, pp. 846–849, 2014.
- [26] X. Wei, N. Chen, and W. Li, "Composite adaptive disturbance observer-based control for a class of nonlinear systems with multisource disturbance," *International Journal of Adaptive Control and Signal Processing*, vol. 27, no. 3, pp. 199–208, 2013.
- [27] H. Sun and L. Guo, "Composite adaptive disturbance observer based control and back-stepping method for nonlinear system with multiple mismatched disturbances," *Journal of the Franklin Institute*, vol. 351, no. 2, pp. 1027–1041, 2014.
- [28] X. Yao and L. Guo, "Composite anti-disturbance control for Markovian jump nonlinear systems via disturbance observer," *Automatica*, vol. 49, no. 8, pp. 2538–2545, 2013.
- [29] X. Wei and N. Chen, "Composite hierarchical anti-disturbance control for nonlinear systems with DOBC and fuzzy control," *International Journal of Robust and Nonlinear Control*, vol. 24, no. 2, pp. 362–373, 2014.
- [30] C. D. Persis, "Robust stabilization of nonlinear systems by quantized and ternary control," *Systems & Control Letters*, vol. 58, no. 8, pp. 602–608, 2008.
- [31] C. De Persis and F. Mazenc, "Stability of quantized time-delay nonlinear systems: a lyapunov-krasowskii-functional approach," *Mathematics of Control, Signals, and Systems*, vol. 21, no. 4, pp. 337–370, 2010.
- [32] T. Liu, Z. P. Jiang, and D. J. Hill, "Small-gain based output-feedback controller design for a class of nonlinear systems with actuator dynamic quantization," *IEEE Transactions on Automatic Control*, vol. 57, no. 5, pp. 1326–1332, 2012.
- [33] S. Tong, S. Sui, and Y. Li, "Fuzzy adaptive output feedback control of MIMO nonlinear systems with partial tracking errors constrained," *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 4, pp. 729–742, 2015.

- [34] Y. Chang, Y. Wang, F. E. Alsaadi, and G. Zong, "Adaptive fuzzy output-feedback tracking control for switched stochastic pure-feedback nonlinear systems," *International Journal of Adaptive Control and Signal Processing*, vol. 33, no. 10, pp. 1567–1582, 2019.
- [35] L. Ma, N. Xu, X. Huo, and X. Zhao, "Adaptive finite-time output-feedback control design for switched pure-feedback nonlinear systems with average dwell time," *Nonlinear Analysis: Hybrid Systems*, vol. 37, Article ID 100908, 2020.
- [36] L. Wang, M. V. Basin, H. Li, and R. Lu, "Observer-based composite adaptive fuzzy control for nonstrict-feedback systems with actuator failures," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 4, pp. 2336–2347, 2018.
- [37] Y.-J. Liu and S. Tong, "Barrier Lyapunov Functions-based adaptive control for a class of nonlinear pure-feedback systems with full state constraints," *Automatica*, vol. 64, pp. 70–75, 2016.
- [38] Y.-J. Liu and S. Tong, "Barrier Lyapunov functions for Nussbaum gain adaptive control of full state constrained nonlinear systems," *Automatica*, vol. 76, pp. 143–152, 2017.
- [39] L. Ma, G. Zong, X. Zhao, and X. Huo, "Observed-based adaptive finite-time tracking control for a class of nonstrict-feedback nonlinear systems with input saturation," *Journal of the Franklin Institute*, vol. 357, no. 16, pp. 11518–11544, 2019.
- [40] J. T. Parker, A. Serrani, S. Yurkovich, M. A. Bolender, and D. B. Doman, "Control-oriented modeling of an air-breathing hypersonic vehicle," *Journal of Guidance, Control, and Dynamics*, vol. 30, no. 3, pp. 856–869, 2007.
- [41] S. Keshmiri, M. Mirmirani, and R. Colgren, "Six-DOF modeling and simulation of a generic hypersonic vehicle for conceptual design studies," in *Proceedings of the AIAA Modeling and Simulation Technologies Conference and Exhibit*, June 2004.
- [42] S. Yin, H. Yu, R. Shahnazi, and A. Haghani, "Fuzzy adaptive tracking control of constrained nonlinear switched stochastic pure-feedback systems," *IEEE Transactions on Cybernetics*, vol. 47, no. 3, pp. 579–588, 2017.
- [43] L. Long and J. Zhao, "Adaptive output-feedback neural control of switched uncertain nonlinear systems with average dwell time," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 7, pp. 1350–1362, 2015.
- [44] B. Niu, H. R. Karimi, H. Wang, and Y. Liu, "Adaptive output-feedback controller design for switched nonlinear stochastic systems with a modified average dwell-time method," *IEEE Transactions on Systems Man & Cybernetics Systems*, vol. 47, no. 7, pp. 1371–1382, 2017.
- [45] D. Shevitz and B. Paden, "Lyapunov stability theory of nonsmooth systems," *IEEE Transactions on Automatic Control*, vol. 39, no. 9, pp. 1910–1914, 1994.
- [46] B. Paden and S. Sastry, "A calculus for computing filippov's differential inclusion with application to the variable structure control of robot manipulators," *IEEE Transactions on Circuits and Systems*, vol. 34, no. 1, pp. 73–82, 1987.

## Research Article

# M&A Short-Term Performance Based on Elman Neural Network Model: Evidence from 2006 to 2019 in China

Ming Xiao <sup>1</sup>, Xionghui Yang <sup>1,2</sup> and Ge Li <sup>1</sup>

<sup>1</sup>School of Economics and Management, University of Science and Technology Beijing, Xueyuan Road No. 30, Haidian District, Beijing 100083, China

<sup>2</sup>Audit Department, CITIC Group Corporation, Guanghua Road No. 10, Chaoyang District, Beijing 100083, China

Correspondence should be addressed to Xionghui Yang; yangxh@citic.com

Received 3 September 2020; Revised 18 November 2020; Accepted 4 December 2020; Published 12 December 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Ming Xiao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Based on the event study method, this paper conducts the analysis on the short-term performance of 1302 major mergers and acquisitions (M&A) in China from 2006 to 2019 and takes the cumulative abnormal return (CAR) as the measurement index. After comparing the five abnormal return (AR) calculation models, it is found that the commonly used market model method and the market adjustment method have statistical defects while the Elman feedback neural network model is capable of good nonlinear prediction ability. The study shows that M&A can create considerable short-term performance for Chinese listed company shareholders. The CAR in window period reached 14.45% with a downward trend, which is the win-win result achieved through the cooperation between multiple parties and individuals driven by their respective rights and interests in the current macro-microeconomic environment in China.

## 1. Introduction

Before 2005, the problem of tradable shares and nontradable shares existed in China's stock market. The controlling shareholders who held nontradable shares were not concerned about the rise and fall of stock price; therefore, the interests of shareholders holding tradable shares cannot be guaranteed. The share-trading reform, launched in 2005 and completed in 2006, made nontradable shares traded, and all shareholders pay more attention to stock prices. 2006 is known as "the year of M&A" [1] because listed companies began to improve the stock price and trading activity through M&A activities.

In 2008, China Securities Regulatory Commission promulgated "Administrative Measures for M&A of Listed Companies," which marked the coming of the era of loose policies of M&A. Since then, a series of policies has been put forward to make the M&A activities more market-oriented.

In the following 10 years, M&A activities of listed companies play an important role in different stages of China's economic development, structural adjustment, transformation, and upgrading. However, the volume and amount of M&A transactions of listed companies increased substantially and attracted the investors to pursue and hype. In 2019, China Securities Regulatory Commission revised "Administrative Measures for M&A of Listed Companies" to strengthen the supervision, prevent arbitrage through M&A, and promote M&A rationality. According to the statistics of Wind, the average amount and number of M&A transactions in China in 2006–2019 were 4.42 trillion Yuan and 5,182, and the number of M&A transactions in the past 3 years exceeded 15,000. M&A has been one of the most important ways of resource allocation for a long time in China's capital market [2].

Since the completion of the share-trading reform in 2006, the discussion about whether M&A can produce

performance and whether it can be used for market value management have never stopped.

## 2. Literature Review

### 2.1. M&A Short-Term Performance Literature

**2.1.1. Synergy Effects Theory.** The synergy effects theory was first proposed by Hermann Haken in 1971 and systematically elaborated in 1976. Since then, it has been applied to the study of M&A motivation theory. According to the efficiency theory proposed by Jensen and Ruback, the important motivation of M&A is that the acquirer and acquiree hope to achieve synergy effects through integration [3], including management, operation, finance, diversification, and other types of synergy [4–7]. M&A gains created by synergy effects will be redistributed among stakeholders, most of which will be transferred to shareholders of both parties during the M&A implementation process [8, 9]. In academic research, the concept of M&A performance is proposed for measuring the synergy effect, and it is divided into long-term performance based on financial index method and short-term performance in view of the event study method [10, 11]. The research object of this paper is the short-term M&A performance of the acquirer. Namely, the CAR on stocks of listed companies in the window period before and after the announcement date is applied as a measure [12].

**2.1.2. Research on Short-Term Performance of Foreign M&A.** The empirical study on the short-term performance of M&A in foreign academics started early (Table 1), and there is no consensus on whether M&A can create short-term performance. Some scholars believe that M&A brings significant positive or negative short-term gains to the acquirer, while others hold that M&A are uncontrollable, which is impossible to bring definite short-term performance to the acquirer.

This paper argues that the inability to reach a consensus conclusion is related to five waves of M&A experienced by Western countries represented by the United States. Scholars have sufficient M&A samples, and the differences in sample scope and time span lead to inconsistent conclusions.

**2.1.3. Research on Short-Term Performance of Chinese M&A.** The empirical research on short-term performance of Chinese M&A started late (Table 2). Due to the speculation and pursue of M&A related stocks by China's stock market for many years, most of the research conclusions focused on the positive short-term performance, and a small number of studies draw different conclusions.

The common feature of short-term M&A performance research in China is that the sample size is small, the coverage period is short, so the sample representativeness and conclusion accuracy are affected, which is related to the objective fact that China's M&A market develops late and the sample of M&A events is small. Zhang empirically analyzed 1,326 M&A events in 1993–2002 based on event study method and concluded that the M&A had a negative

impact on the acquirer with  $-16.76\%$  CAR during the window period. This literature is rare M&A performance research based on larger sample sizes. However, all the event samples occurred before 2005's share-trading reform, and most of them have no exact M&A announcement date. Chen et al. (2017) found that the share-trading reform had a positive impact on China's M&A performance [34], because the improvement of stock liquidity enhanced the reaction speed of stock prices to major decisions of company managers [35]. Therefore, it is necessary to make further researches on the short-term performance of China's M&A after share-trading reform.

### 2.2. Literatures of M&A Short-Term Performance Measurement

**2.2.1. Event Study Method.** Bruner proposed four measurement methods of M&A performance as follows: the event study method, the financial index method, the case-study method, and the management personnel interview method. Among those, the event study method is one of the most important methods for scholars to study M&A performance. The event study method is a general term for a series of methods for measuring the degree of influence of an event on the price of a particular financial asset [36] and has been widely accepted by scholars after improved by Ball and Brown [37] and Fama et al. in the study of market effectiveness [38].

The calculation of the impact of M&A events on stock prices by abnormal return (AR) has become the mainstream method for M&A performance research at home and abroad [39]. AR refers to the return difference between the stock's actual return rate and the normal (predicted) return rate under the assumption without the M&A transactions. The primary task in the calculation of AR is how to design a model to predict normal return.

**2.2.2. Algorithms of AR.** The AR algorithms commonly applied by scholars include the market adjustment method and the market model method [40]. The former assumes that the normal return is the market index return rate, while the latter calculates the normal return based on the capital asset pricing model (CAPM). In addition to utilizing the above traditional methods, foreign scholars have tried other methods to improve the accuracy of AR. For instance, Gregory [41] adopted the market model method, risk- and size-adjusted model, simple size-adjusted model, and value-weighted three-factor model proposed by Fama and French to calculate AR in M&A, finding significant differences in different AR algorithms. Besides, in the study of 1,164 M&A events in the United States from 1955 to 1987, Agrawal et al. [15] utilized the AR algorithms from the studies of Dimson and Marsh [42], Lakonishok and Vermaelen [43], and Ibbotson [44].

Almost all of Chinese scholars adopt the market adjustment method or the market model method, lacking the attempt and exploration of the AR algorithm. Zhang compared the market adjustment method with the market



TABLE 1: Overview of short-term performance research on foreign M&amp;A.

| Conclusion category             | Author                     | Time published | Number of samples | Sample year | Research conclusion  |
|---------------------------------|----------------------------|----------------|-------------------|-------------|--|
| Positive short-term performance | Madden Gerald              | 1981           | 86                | 1997–1979   | One day before and after the M&A announcement, the CAR was significantly positive, but when the window period was extended, the CAR was significantly reduced [13]                                   |
|                                 | Healy and Palepu           | 1992           | 50                | 1979–1984   | Enterprises with highly similar products obtained more positive M&A performance and are the acquirers' capital productivity was raised significantly [14]  |
|                                 | Agrawal et al.             | 1992           | 1,164             | 1955–1987   | About half of the acquirer shareholders were able to obtain a positive CAR, and the CAR gradually decreased with the extension of the window period [15]   |
|                                 | Humphery-Jenner and Powell | 2014           | 17,647            | 1996–2008   | M&A samples from 45 countries showed that the acquirer created positive short-term performance, which decreased with the increasing national governance intensity [16]                               |
| Negative short-term performance | Dodd                       | 1980           | 172               | 1973–1976   | The CAR of the acquirer in about half of the sample during the 2-day window period before and after the M&A announcement was significantly negative [17]   |
|                                 | Higson and Elliott         | 1998           | 830               | 1975–1990   | The CAR of the acquirer in window period was significantly negative [18]   |
|                                 | Hans                       | 2006           | 110               | 1993–2001   | M&A not only brought negative cumulative returns to the acquirer but also continued to decline as the window period increased [19]   |
| Uncontrollability               | Jarrell.                   | 1988           | 663               | 1962–1985   | According to the time of M&A announcement, the samples were divided into three groups; the CAR of the acquirer was inconsistent among the three groups, and there was no significant difference [20] |
|                                 | Bruner                     | 2002           | N/A               | 1971–2001   | After the summary of 130 classics from 1971 to 2001, it was concluded that there was uncertainty in the short-term M&A performance [21]  |
|                                 | Yook                       | 2004           | 75                | 1989–1993   | The short-term performance for acquirer was not significant due to the influence of the premium of the acquisition target [22]   |
|                                 | Uddin                      | 2009           | 373               | 1994–2003   | M&A did not bring significant short-term performance to the acquirer [23]  |

TABLE 2: Overview of short-term performance research on Chinese M&amp;A.

| Conclusion category             | Author          | Time published | Number of samples | Sample year | Research conclusion  |
|---------------------------------|-----------------|----------------|-------------------|-------------|--|
| Positive short-term performance | Li and Chen     | 2002           | 349               | 1999–2000   | M&A brought significant wealth increase to the acquirer shareholders, especially the acquirer shareholders with larger proportion of national or legal person shares [24].       |
|                                 | Liu et al.      | 2009           | 749               | 1998–2004   | During the window period, acquirer shareholders received an average of 1.39% CAR, explaining the conclusion based on the industry cycle theory [25]                              |
|                                 | Deng et al.     | 2011           | 312               | 1997–2000   | Non-associated M&A created significant returns for the acquirers, and the associated M&A did not create wealth for shareholders [26]   |
|                                 | Zhang and Sheng | 2016           | 55                | 2010–2016   | M&A in the Internet finance industry brought significant positive short-term performance, and mixed M&A performance was better than horizontal and vertical M&A performance [27] |
|                                 | Li and Song     | 2017           | 333               | 2010–2013   | M&A created significant short-term M&A gain and increased as risk investor participation grows [28]  |

TABLE 2: Continued.

| Conclusion category             | Author         | Time published | Number of samples | Sample year | Research conclusion   |
|---------------------------------|----------------|----------------|-------------------|-------------|---|
| Negative short-term performance | Zhang and Lei  | 2003           | 216               | 1999–2001   | The wealth of the acquirer shareholders did not increase due to M&A activity, and the CAR rose first and then decreased, and the reduction was greater than the increase [29]                             |
|                                 | Zhang          | 2003           | 1,326             | 1993–2002   | M&A had a negative impact on the acquirer with $-16.76\%$ CAR during the window period [12]   |
|                                 | Zhu and Chen   | 2016           | 517               | 2011–2013   | Technology M&A brought significant negative short-term performance to the acquirer, but the company's establishment period and equity concentration were conducive to improving M&A performance [30]      |
| Uncontrollability               | Chen and Zhang | 1999           | 95                | 1997        | Due to the immature capital market in China, the main M&A stocks did not show significant fluctuations, and the stock market did not respond significantly to M&A [31]                                    |
|                                 | Yu and Yang    | 2000           | 18                | 1993–1995   | In the M&A, the enterprise value of the acquirer did not rise, and the shareholders were not able to obtain returns, which did not benefit the development of the enterprise [32]                         |
|                                 | Yu and Liu     | 2004           | 55                | 2002        | The M&A performance of the acquirer was not significant, and lacked continuity. From the perspectives of M&A motives and methods, the causes for the high failure rate of M&A in China were analyzed [33] |

model method, finding the same conclusion in measuring M&A performance. In addition to adopting the above two methods, Cong made an attempted to utilize the listed company's return on net assets per share to minus the market interest rate to calculate the AR [45], which has become a rare Chinese literature on the AR algorithm research.

The market adjustment method and the market model method have distinct advantages and disadvantages. The former is simple in calculation but lacks theoretical and statistical basis. The latter possesses theoretical basis, but its hypothesis testing results were rarely discussed systematically in previous literatures. The empirical study of this paper shows that the regression equation coefficients cannot pass the significance  $T$  test, which is related to the nonlinear characteristics of the stock price series and is ignored usually due to the passing of  $F$  test with the regression equation. In order to solve the nonlinear problem, this paper designs another two traditional regression models and one artificial intelligence model to calculate AR and compares the fitting effect, prediction accuracy, and significant difference between the five models, which fills in the literature blank about AR algorithm. The accuracy of AR calculation is the basis of all the literatures on the application of event study method, which directly affects the results of events. Therefore, it is necessary and meaningful to carry out the research of AR algorithm.

**2.2.3. Artificial Neural Network Algorithm.** With the continuous improvement of chaos and fractal theory, considerable studies have proved that the stock price series owns nonlinear characteristics [46–50], and scholars have begun to utilize some data mining techniques to solve complex nonlinear problems [51]. Artificial neural network (ANN) is

an adaptive nonlinear dynamic system composed of a large number of neurons through extremely flexible and extensive connections [52], with self-learning, self-organization, and self-adaption functions, which can reveal the complexities contained in data samples [53, 54], and has been widely applied in financial time series studies since the 1990s [55]. Moreover, it has proven to be more suitable for stock forecasting than traditional linear models [56]. Ican and Çelik [57] compared 25 literatures based on neural network predicting stock prices, holding that selecting the appropriate stock data (input information) and neural network structure have an important influence on the fitting effect. According to the topology of neuron connections, neural networks can be divided into forward networks (such as BP neural networks) and feedback networks (such as Elman neural networks). In contrast to forward network, feedback networks can achieve information feedback and have associative memory functions. Weng and Lin [58] compared the short-term prediction effects from stock prices of three neural networks (RBF, BP, and Elman). The empirical results revealed that the prediction ability of Elman feedback neural network was higher than that of the other two forward neural networks.

The core idea of Elman feedback neural network originates from the simple recurrent neural network model proposed by Jeffrey Locke Elman in 1990, consisting of input layer ( $L1$ ), hidden layer ( $L2$ ), connection layer ( $L3$ ), and output layer ( $L4$ ) (Figure 1), which is frequently applied for dynamic modeling or time series prediction [59]. The input information ( $XN$ ) enters the hidden layer neurons through the input layer neurons, and the output information of the hidden layer is calculated and stored by the connected layer neurons and then enters the hidden layer as input information again, repeating iteratively until the error function and the weight reach a stable balance state (Figure 2).

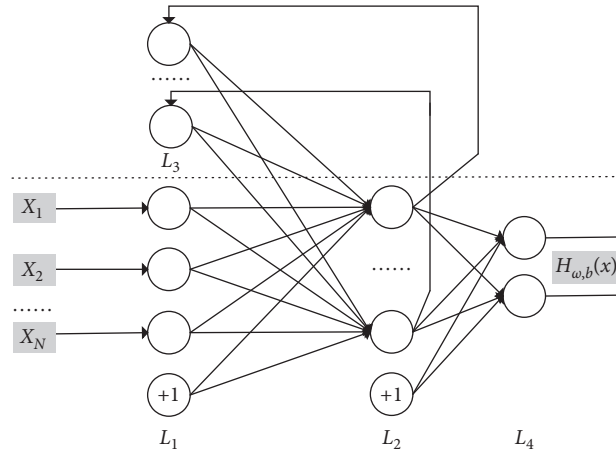


FIGURE 1: Elman neural network structure diagram.

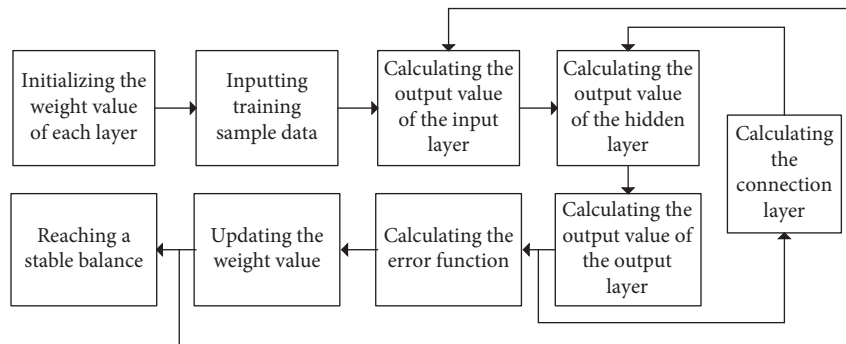


FIGURE 2: Elman neural network operation logic.

Among these, the activation functions of the output layer and the connection layer are linear functions, and the activation function of the hidden layer is a nonlinear function [60].

Since the beginning of this century, Elman neural network has been widely applied in the research on stock trading strategy and trading timing. Sitte and Sitte demonstrated that the S&P500 index can be predicted, through applying the Elman neural network [61]; Huang et al. utilized the Elman neural network to forecast the direction of the stock market and achieved better predictions [62]. Hyun and Kyung introduced the idea of the genetic algorithm based on Elman neural network for financial time series prediction, and the prediction accuracy was further improved [63]. Chinese research on Elman neural network for stock forecasting started late, and scholars have modified the structure or parameters of Elman neural network to study different financial time series predictions. It is agreed that Elman neural network has better nonlinear prediction ability [64–66].

### 3. Data and Methodology

**3.1. The Innovation of This Paper.** Scholars inside and outside China mainly adopt the market adjustment method and market model method to calculate AR. This paper

applies another three models, including the Elman neural network model, to compare and improve the rigor and accuracy of AR calculation, which is the first academic attempt. In addition, this paper takes the Chinese share-trading reform as the starting point and selects almost all M&A events with trading suspension and resumption as research sample to study the changes of M&A performance over the past 14 years, which makes up for the shortcomings of the small coverage period of Chinese M&A samples.

**3.2. Sample Selection.** This paper collects 2,358 major M&A events of listed companies from 2006 to 2019 from the Wind M&A database. The remaining 1,302 M&A events are the total sample, after eliminating 1,056 events failed, unfinished, or in which listed companies as acquiree, or no exact M&A announcement date due to the small transaction volume.

**3.3. Research Model.** The short-term M&A performance indicator adopts the cumulative abnormal return (CAR) of the window period, which is 21 days around M&A announcement, marked as  $(-10, 10)$  with 0 being the announcement day. In this paper, five models are applied to predict the normal return and then calculate the AR. The merits and demerits are compared by three factors: the

determinable coefficient ( $R^2$ ), the root mean-squared error (RMSE), and the significant difference test. The model design adopts Matlab math software [67].

**3.3.1. Market Adjustment Method.** Under the market adjustment method, it is not necessary to determine the observation period. The market index yield is directly subtracted from the stock actual return to calculate the AR. Due to its simple calculation, it is widely applied. As a matter of fact, this model is to assume the constant term and the risk coefficient in the CAPM model as 0 and 1, respectively. This assumption is neither theoretical nor consistent with the reality.

**3.3.2. Market Model Method.** The market model method equation is given as follows:

$$\hat{R}_i = \beta_{im} * R_m + \alpha + \varepsilon. \quad (1)$$

The market model method is a unary linear model based on CAPM theory (equation (1)). It is necessary to predict the constant term  $\alpha$  and the risk coefficient  $\beta_{im}$  according to the linear relationship between stock return and the market yield in the observation period. This paper makes the improvements as follows: first, the observation period of each M&A event is selected by finding trading day range with highest correlation coefficient between the stock return and the market yield before the window period, so as to improve the goodness of fit. The average value of the highest correlation coefficient of all samples in the observation period is 0.6444. If the observation period is set as fixed interval of 50 days before the window period, the average value is 0.5566, which shows that the linear relationship in the observation period is significantly improved. Secondly, on the basis of the first fitting, the noise outliers outside the two standard deviations near the fitted line are eliminated (Figure 3(d)), and then the second fitting is performed. The confidence interval (Figure 3(b)) after eliminating abnormal value is more concentrated than before (Figure 3(a)).

After excluding the outliers, there are only 224 events whose constant term  $\alpha$  and risk coefficient  $\beta_{im}$  both pass the significance test (0.05), and the average coefficient  $R^2$  is 0.5479, indicating that the explanation and prediction ability of market yield to stock return is weak under the unitary linear model, which is consistent with the doubts about the CAPM theory in the previous literature [68]. In this case, this paper attempts the unary nonlinear model.

**3.3.3. Unary Nonlinear Model.** The unary nonlinear model equation is given as follows:

$$y_i = \beta_1 e^{(\beta_2 / (x_i + \beta_3))} + \varepsilon_i, \quad \varepsilon_i \sim N(0, \delta^2). \quad (2)$$

Considering that the distribution of AR is dense with a large fluctuation, the negative exponential function (equation (2)) with the trend of steepness first and then slowness is selected as the unary nonlinear regression model. This paper makes the improvements as follows: First, in order to cover

the stock return history as much as possible and avoid the long-term observation period to damage the goodness of fitting, we take 5 trading days as the step value for each event. From 30 days before the window period, the observation period will be gradually expanded forward for fitting. The observation period with the smallest root mean-squared error (RMSE) is selected as the optimal observation period, and the average observation period of all samples is 51 days. Secondly, the second time fitting is eliminated on the basis of the first fitting, and the improvement effect of the goodness of fit is significant (Figure 4).

After the elimination of the outliers, there are only 83 events whose all the three parameters  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  pass the significance test (0.05), and the determination coefficient  $R^2$  is 0.5472, which is mainly due to the fact that it is difficult to predict the specific analytical formulae of the nonlinear relationship in practice. In this case, a one-dimensional polynomial model can be tried to gradually fit the measured points.

**3.3.4. Unary Polynomial Model.** Any function can theoretically be approximated by a polynomial model by segmentation (equation (3)). Hence, this paper is fitted from low order to high order, and the improvements are made as follows: first, each observation period of each M&A event is performed to fit from the first order to the tenth order. Secondly, 80 kinds of observation period are selected for each M&A event, which are 21 days, 22 days . . . 100 days before the window period. Each event is fitted for 800 times based on the 1–10 order and 80 observation periods, and the equation with the smallest RMSE is selected as the optimal order and the optimal observation period:

$$y_i = p_1 x_i^n + p_2 x_i^{n-1} + \dots + p_n x_i + p_{n+1} + \varepsilon_i, \quad \varepsilon_i \sim N(0, \delta^2). \quad (3)$$

The optimal order of all samples is 10, the optimal observation period is 33 days, and the average coefficient of  $R^2$  is 0.6790. The fitting effect is greatly improved. However, when the model is utilized to predict AR in the window period, the unreasonable extreme value accounts for 16.70%. The high-order polynomial model can only fit the limited data in observation period, and when data that cannot be covered in observation period appear in window period, the amplification function of the high-order items in the model will destroy the prediction ability. Figure 5(a) illustrates a better approximation fitting of the fitted curve to the limited data in the observation period. When the Y-axis display range is expanded to show the overall trend of the fitted curve, Figure 5(b) reflects the excessive fluctuation characteristics of the high-order fitting curve. Therefore, when the order is limited to the 4th order or less, the extreme value is basically eliminated. However, the coefficient  $R^2$  is reduced to 0.0163.

**3.3.5. Elman Neural Network Model.** The purpose of traditional regression analysis is to find the mapping relationship between independent variables and dependent

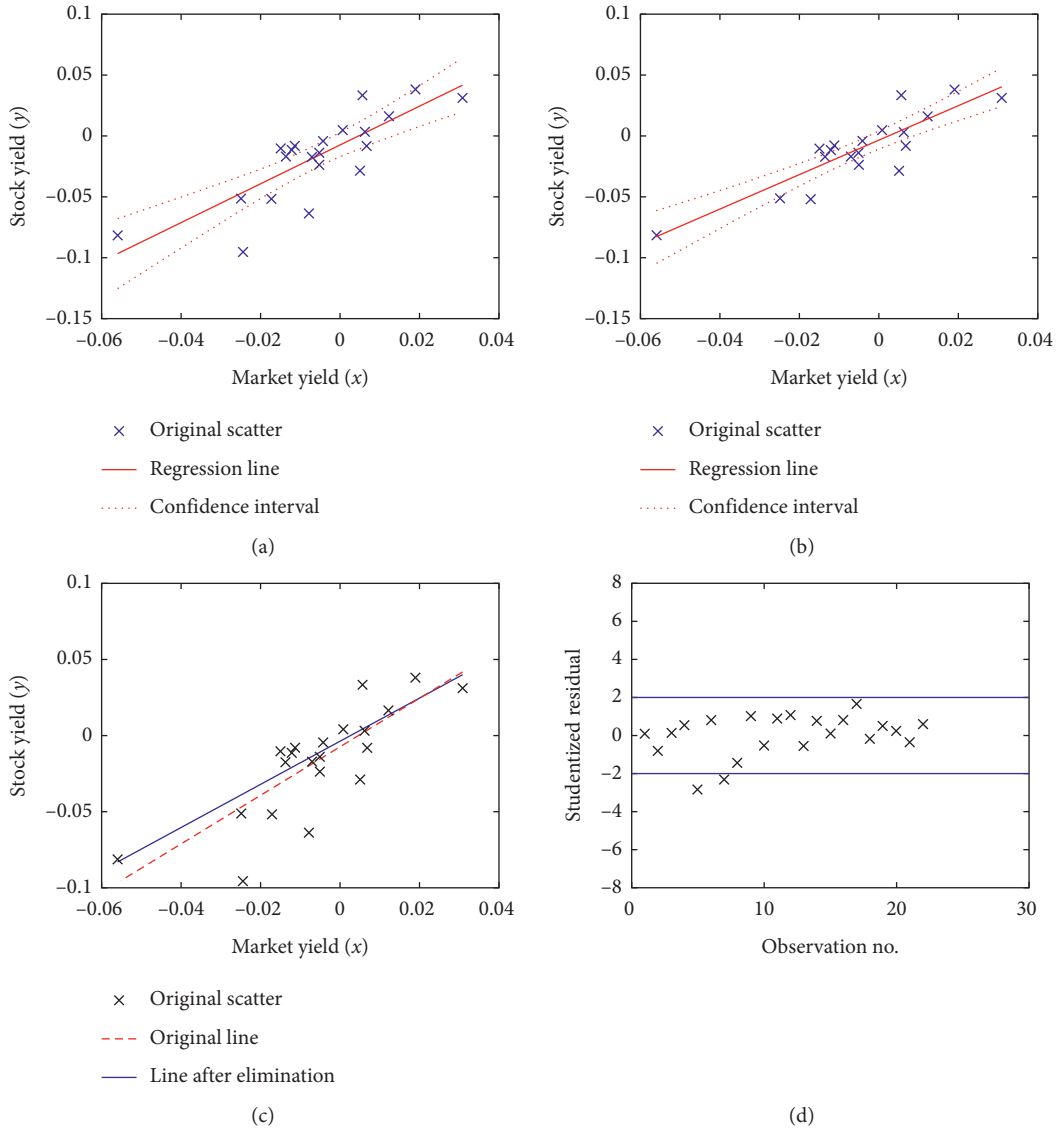


FIGURE 3: Market model method fitting map. (a) Before eliminating abnormal value. (b) After eliminating abnormal value. (c) Before and after eliminating abnormal value. (d) Studentized residual map.

variables. The results of the above four models show that it is difficult to find analytical expressions that satisfy both the hypothesis test condition and the predictive ability in practice. The main cause is that the complex relationship in financial time series is difficult to determine with the function of the analytical expression. As one of the data mining techniques, Elman neural network is widely applied in autonomous learning, associative storage, and high-speed optimization. Theoretically, it can handle arbitrary complex causal relationships, which is suitable for stock return forecasting.

The improvements are made in the Elman model as follows: first, the stock normal return is predicted by input information with individual stock's historical returns ( $E1$ ) and market yields ( $E2$ ), respectively. Secondly, the Elman model memory function is fully applied to cover the stock return history as much as possible with the observation

period selected from 2 months after listing to before the window period. The average observation period of all samples is 1,940 days; the maximum number of iterations is 2,000, and the error tolerance is 0.00001. The iteration process is stopped when the mean-squared error (MSE) reaches the error tolerance. If the error tolerance is not reached after 2,000 iterations, then the parameters, such as the weight and activation function corresponding to the minimum MSE, are taken as the optimal solution. Figure 6 illustrates the process of reducing the MSE to 0.0005 after 2,000 iterations in one of the M&A events.

The  $R^2$  of the Elman model's fitting with individual stock return or market yield is 0.9859 and 0.9958, respectively. The latter is better than the former. This conclusion can also be obtained from the fluctuation of the residual plot (Figures 7 and 8), indicating that the linkage between individual stock return and the market yield is stronger than that between

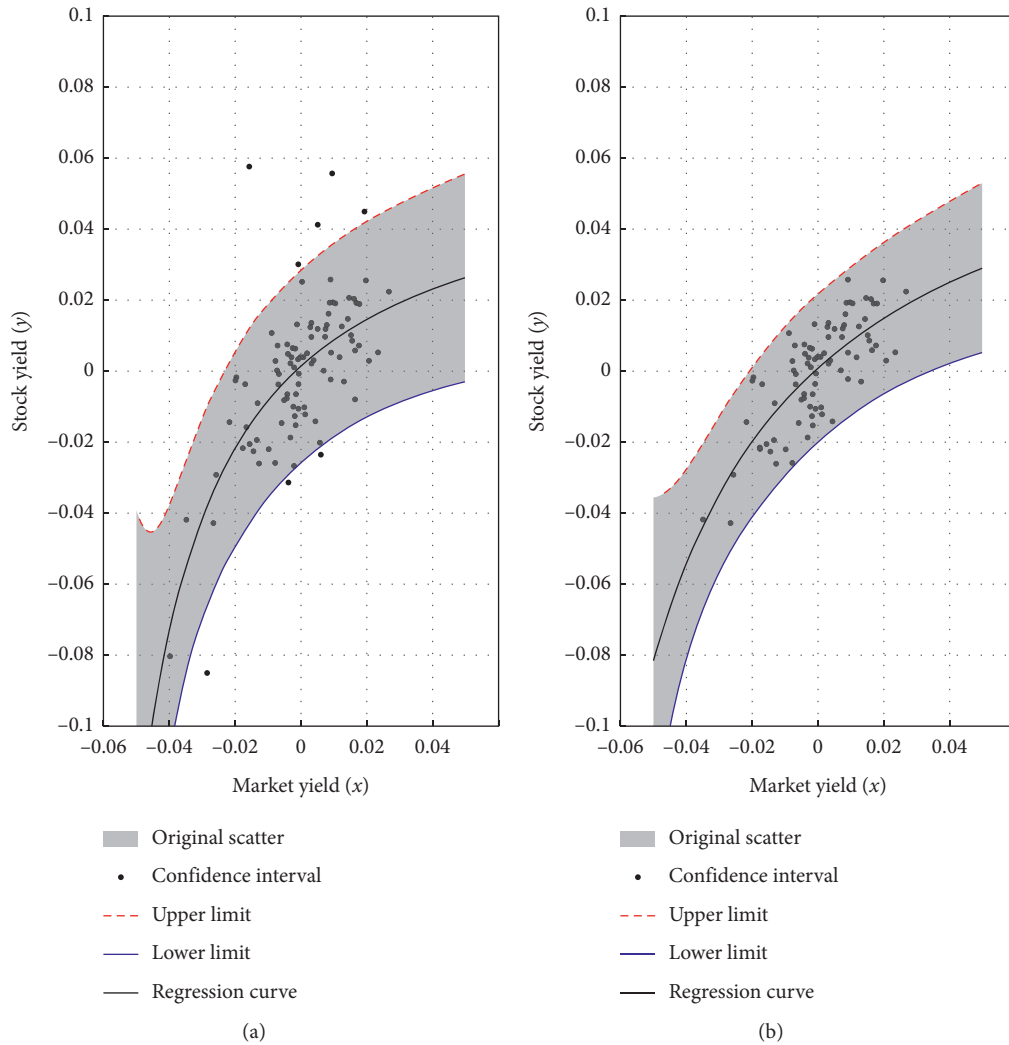


FIGURE 4: Unary nonlinear model fitting map. (a) Before eliminating abnormal value. (b) After eliminating abnormal value.

individual stock return and their own historical return. The fitting effect is shown in Figure 9.

#### 4. Results and Discussions

The five models have different window yields (Table 3). When studying the AR in window period, the following important issues are rarely demonstrated or mentioned: (1) whether the yield rate during the observation period meets the assumptions of the classical theory and the regression model; (2) whether the mean of the AR in the window period is representative; (3) whether there is a significant difference between the window period's AR time series calculated by the five models; (4) whether the overall trend of the AR in each year is significant.

*4.1. Normality Test of the Yield Rate in Observation Period.* The financial time series is supposed to obey the normal distribution, which is almost the common assumption of all classical theories (such as CAPM theory) and the traditional regression model because the normal distribution possesses

good additivity. Based on the central limit theorem, the totality can be considered to obey the normal distribution when the sample size is greater than 30. In this case, the hypothesis that the stock yield rate series obeys a normal distribution is widely dictated with its verification ignored. Considering the advantages as well as disadvantages of various methods, this paper utilizes six common normal distribution test methods (Table 4).

The data period of yield rate series was taken 50 days before the window period. The normality test was carried out for each individual stock and the market yield rate series. The mean skewness of individual stock and the market was  $-0.0242$  and  $-0.8370$ , respectively. The mean kurtosis was  $4.6782$  and  $6.6067$  respectively, indicating that the sequence has a peak fat tail characteristic, which was verified through utilizing the other four methods. Although the results of JB,  $\chi^2$ , and Lilliefors are slightly different, the conclusions are basically the same. Namely, the yield series of large proportion (up to 45.16%) does not obey the normal distribution, and the assumptions of the classical theory and the regression model are not true. Therefore, the statistical

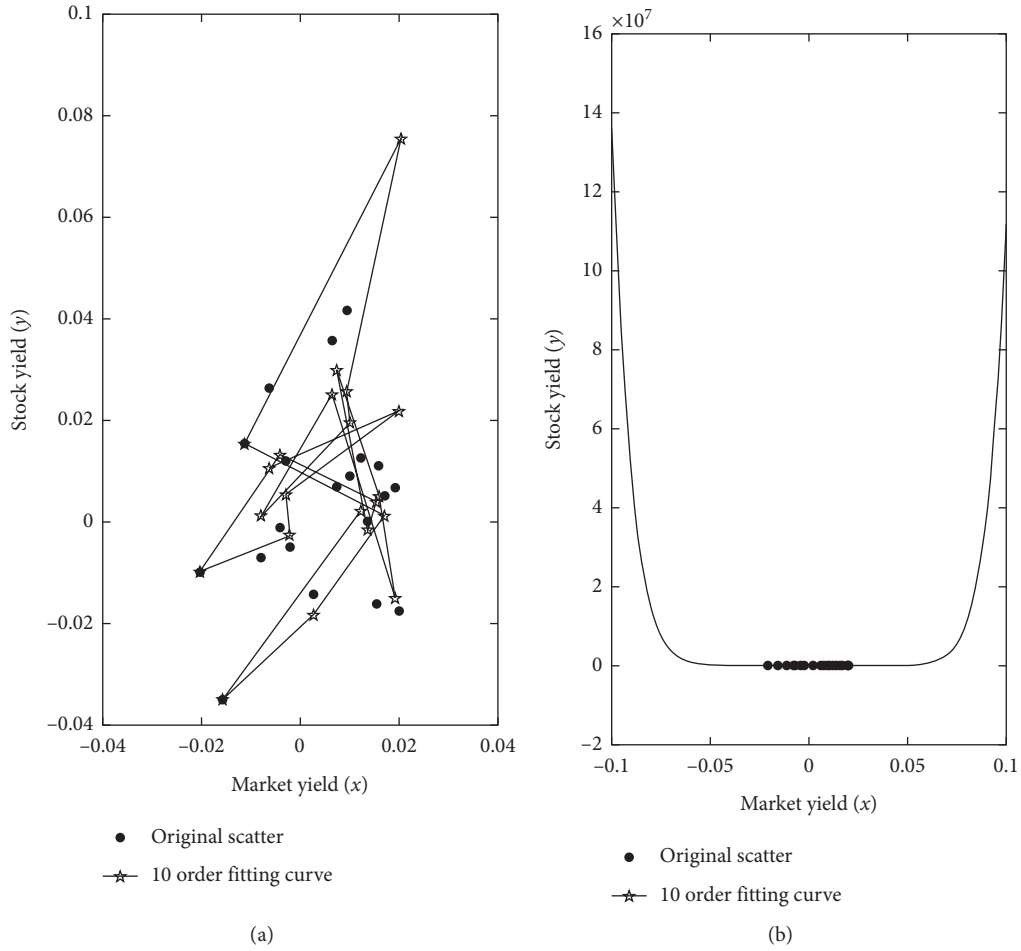


FIGURE 5: High-order polynomial model fitting map. (a) Original scatter VS predicted points. (b) High-order polynomial fittings.

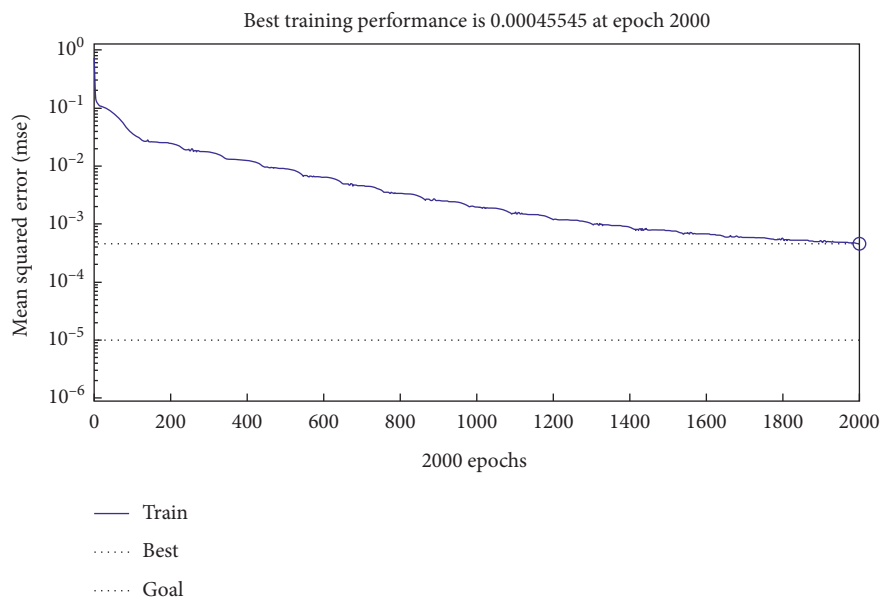


FIGURE 6: Elman iterative MSE convergence graph.

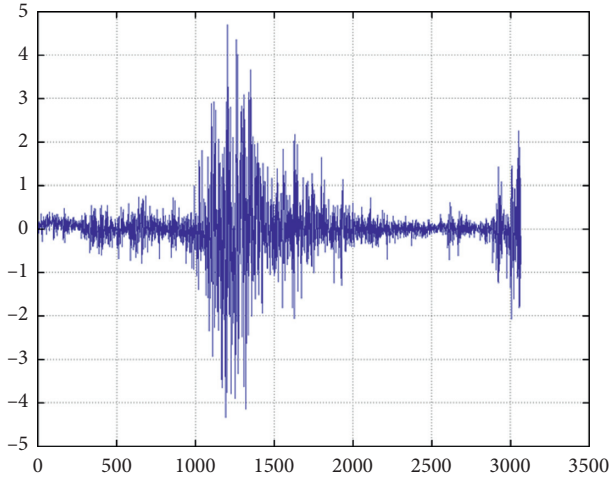
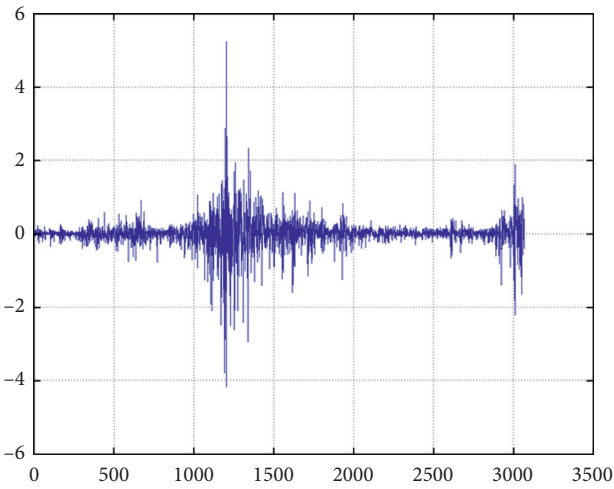
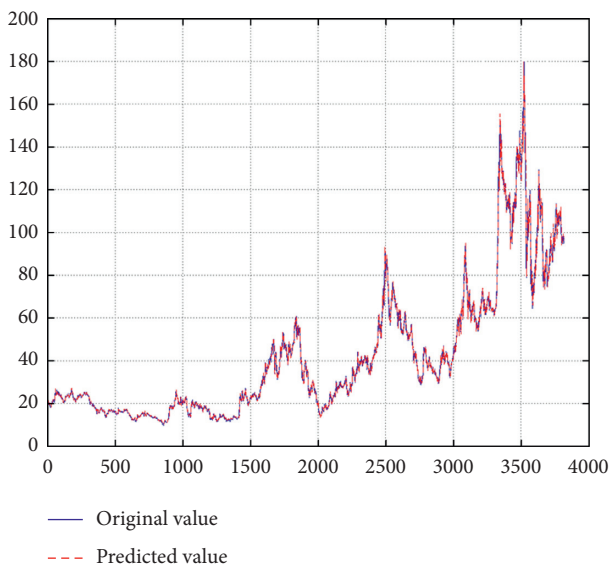
FIGURE 7: Elman residual plot under  $E1$ .FIGURE 8: Elman residual plot under  $E2$ .

FIGURE 9: Elman model fitting map.

model without normality requirements for the data should be selected and applied.

The K-S test results are quite different from other methods because this paper replaces the population parameters with sample mean and standard deviation like other related literature practices. In essence, it has change the K-S test to Lilliefors test. However, the statistical software (such as Matlab and SPSS) defaults and utilizes the K-S test threshold table, and the mismatch between Lilliefors statistic and K-S test threshold table results in an incorrect conclusion [69].

*4.2. Representative Test for Abnormal Return Mean.* After the representative test, for each model, all the average ARs of every day in the window period (the abnormal return in Table 3) pass the significance test of the 0.005 level.

*4.3. Significant Difference Test between the AR of the 5 Models.* Although the average daily ARs in the window period under the five models are representative, the comparison between average daily ARs cannot infer whether there is a significant difference between AR series under different models. Since AR series do not have normality and homogeneity of variance, the Friedman test (two-factor rank variance analysis) of the nonparametric test method is applied to demonstrate significant differences between AR series of any two models on the same day.

The Friedman test results (Table 5) can be concluded in three aspects as follows: (1) there are 17 days(80.95%) with significant differences in 21 days' window period in AR series between the market adjustment method and the market model method, which means that the substitution of the former for the latter in some literatures is not rigorous and affects accuracy of AR. (2) Because the low-order polynomial method is too gentle and can only reflect the overall trend, there are 0 day with significant differences with the market adjustment method. (3) There are 20 or 21 days' significant difference between Elman model and other four models, which means Elman model is totally different from other models.

*4.4. Significance Test of AR Series' Change in Window Period under Elman Model.* The Friedman test is carried out on the significant difference of AR series in two adjacent days in the window period under the Elman model. The results show that there are significant differences in AR series between 2 days before and 3 days after the announcement date. In Figure 10, the 21 horizontal lines represent the rank and mean confidence intervals (0.05) of the daily AR series, and the vertical axis 11 represents the announcement day. There is no overlap in the projection of adjacent horizontal lines of the ordinate 9–14 in horizontal axis, indicating a significant difference in AR series between the two adjacent days.

The settlement results of the Elman model is shown as follows: (1) M&A news has been transmitted to the stock market at least 2 days before the announcement date, which causes the stock price to fluctuate significantly within 6 days



TABLE 3: The AR in window period and fitting result of the five models.

| Window period  | Actual return (%) | Abnormal return              |                         |                           |                              |           |
|----------------|-------------------|------------------------------|-------------------------|---------------------------|------------------------------|-----------|
|                |                   | Market adjustment method (%) | Market model method (%) | Unary nonlinear model (%) | Unary polynomial model I (%) | Elman (%) |
| -10            | 0.20              | 0.05                         | 0.11                    | -0.01                     | -0.02                        | 0.01      |
| -9             | 0.43              | 0.26                         | 0.30                    | 0.44                      | 0.15                         | 0.14      |
| -8             | 0.26              | 0.25                         | 0.33                    | 0.52                      | 0.08                         | 0.17      |
| -7             | 0.07              | 0.09                         | 0.15                    | 0.28                      | -0.15                        | -0.04     |
| -6             | 0.35              | 0.17                         | 0.21                    | 0.35                      | 0.05                         | 0.05      |
| -5             | 0.15              | 0.20                         | 0.33                    | 0.10                      | 0.03                         | 0.03      |
| -4             | 0.32              | 0.18                         | 0.27                    | -0.40                     | 0.08                         | 0.04      |
| -3             | 0.36              | 0.35                         | 0.43                    | 0.56                      | 0.20                         | 0.27      |
| -2             | 0.36              | 0.33                         | 0.45                    | 0.42                      | 0.10                         | 0.30      |
| -1             | 1.39              | 1.19                         | 1.27                    | 1.36                      | 1.11                         | 1.15      |
| 0              | 4.15              | 3.94                         | 4.00                    | 1.27                      | 3.80                         | 3.86      |
| 1              | 3.14              | 3.14                         | 3.30                    | 3.36                      | 2.91                         | 3.08      |
| 2              | 2.46              | 2.25                         | 2.35                    | 2.50                      | 2.14                         | 2.06      |
| 3              | 1.52              | 1.37                         | 1.42                    | 1.60                      | 1.25                         | 1.15      |
| 4              | 0.89              | 0.79                         | 0.86                    | 1.07                      | 0.60                         | 0.64      |
| 5              | 0.76              | 0.65                         | 0.75                    | 0.08                      | 0.38                         | 0.59      |
| 6              | 0.62              | 0.50                         | 0.58                    | 0.71                      | 0.31                         | 0.49      |
| 7              | 0.08              | 0.19                         | 0.28                    | 0.34                      | -0.04                        | 0.13      |
| 8              | 0.33              | 0.23                         | 0.34                    | 0.26                      | 0.01                         | 0.26      |
| 9              | 0.13              | 0.11                         | 0.22                    | 0.05                      | -0.10                        | 0.08      |
| 10             | 0.10              | 0.05                         | 0.11                    | -0.21                     | -0.06                        | -0.01     |
| CAR            | 18.07             | 16.28                        | 18.03                   | 14.65                     | 12.84                        | 14.45     |
| R <sup>2</sup> | —                 | 0.2813                       | 0.5346                  | 0.5459                    | 0.0163                       | 0.9950    |

TABLE 4: Advantages and disadvantages of the normality test method and the events failed.

| Testing method                | Characteristics  | Individual stock | Market       |
|-------------------------------|--|------------------|--------------|
| Skewness test (mean)          | Simple but not comprehensive, susceptible to extreme peak values   | -0.0242          | -0.8370      |
| Kurtosis test (mean)          | Simple but not comprehensive, susceptible to extreme peak values   | 4.6782           | 6.6067       |
| J-B test                      | Susceptible to outliers based on skewness and kurtosis   | 493 (37.86%)     | 588 (45.16%) |
| $\chi^2$ goodness of fit test | First grouping and posttesting, suitable for category data, easy to make false errors                      | 275 (21.12%)     | 302 (23.20%) |
| Lilliefors test               | Suitable for the unknown overall parameter, applying the sample statistic instead of the overall parameter | 376 (28.88%)     | 588 (45.16%) |
| K-S test                      | Suitable for continuous quantitative data with units of measurement and test of full observation points    | 11 (1.90%)       | 14 (1.08%)   |

TABLE 5: The days with significant difference in the window period between 5 models.

| Model combination        | Market adjustment | Market model | Unary nonlinearity | Polynomial (low-order) | Elman       |
|--------------------------|-------------------|--------------|--------------------|------------------------|-------------|
| Market adjustment method | —                 | 17 (80.95%)  | 9 (42.86%)         | 0                      | 21 (100%)   |
| Market model             | 17 (80.95%)       | —            | 7 (33.33%)         | 13 (61.90%)            | 21 (100%)   |
| Unary nonlinearity       | 9 (42.86%)        | 7 (33.33%)   | —                  | 11 (52.38%)            | 20 (95.24%) |
| Polynomial (low-order)   | 0                 | 13 (61.90%)  | 11 (52.38%)        | —                      | 20 (95.24%) |
| Elman                    | 21 (100%)         | 21 (100%)    | 20 (95.24%)        | 20 (95.24%)            | —           |

around the announcement date. After that, the AR gradually decreases and creates a CAR of 14.45% in the window period; (2) after the announcement day, the AR, which is

significantly different from the previous day, can still continue for 4 consecutive days. The impact of M&A disclosure on the stock market does not disappear immediately, and

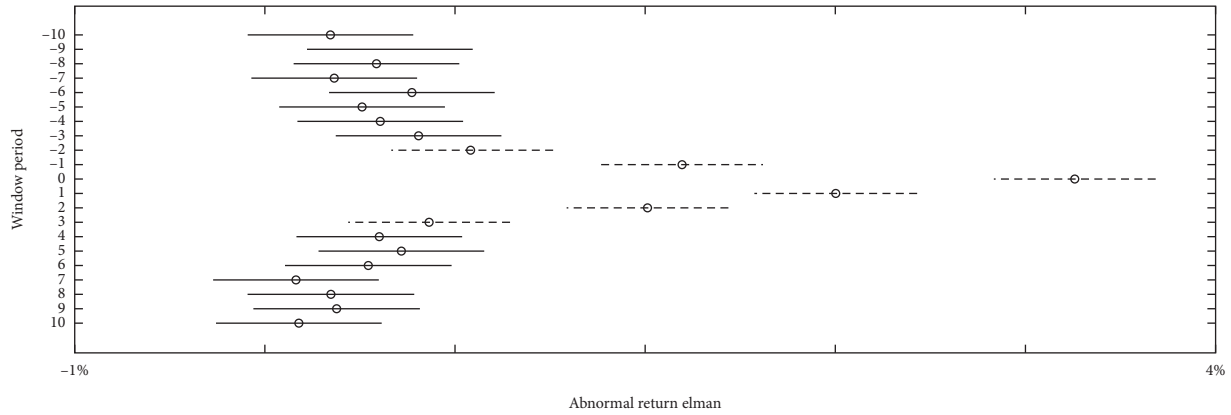


FIGURE 10: Significance test change in the window period.

TABLE 6: Annual distribution of research samples (1302) and cumulative average abnormal return.

| Year                         | 2006–2008 | 2009  | 2010  | 2011  | 2012 | 2013  | 2014  | 2015  | 2016 | 2017  | 2018  | 2019  | 06–19 |
|------------------------------|-----------|-------|-------|-------|------|-------|-------|-------|------|-------|-------|-------|-------|
| Number                       | 18        | 25    | 32    | 34    | 56   | 99    | 183   | 299   | 237  | 151   | 131   | 37    | 1302  |
| CAR (%)                      | 45.21     | 31.02 | 26.90 | 13.91 | 8.88 | 24.30 | 26.18 | 28.25 | 8.84 | -3.51 | -3.45 | 12.36 | 14.45 |
| Announcement day (%)         | 5.49      | 6.91  | 7.15  | 4.03  | 3.54 | 7.86  | 7.43  | 4.39  | 3.12 | 0.15  | -0.27 | 3.84  | 3.86  |
| Next day of announcement (%) | 4.36      | 2.67  | 1.59  | 1.99  | 1.07 | 1.35  | 1.02  | 0.86  | 1.59 | 0.74  | 0.64  | 3.05  | 3.08  |

TABLE 7: Comparison of advantages and disadvantages of five AR models.

| Model              | Market adjustment method  | Market model  | Unary nonlinearity   | Unary polynomial  | Elman model  |
|--------------------|---------------------------|---|--|---|--|
| Advantage          | Easy calculation          | Simple calculation with theoretical basis   | Simple calculation   | Infinite fitting can be realized theoretically by increasing the order  | Can solve complicated nonlinear causality problem  |
| Disadvantage       | Lack of theoretical basis | Difficult to pass equation parameters' significance test                                      | Difficult to pass equation parameters' significance test                                 | Low-order: Poor fitting; high-order: Poor predictive ability  | The program is complicated; the output layer information is underutilized  |
| Innovation attempt | None                      | Take maximum correlation coefficient's interval as observation period; eliminate the outliers | Selecting the best observation period by step progressive method; eliminate the outliers | Selecting the best observation period by the step progressive method; comparison of low-order and high-order fitting/prediction results | Keep the observation period as much as possible; comparison of closing price and yield rate as input information |

China's stock market has not yet reached the semistrong position according to the semistrong effective judgment standard.

4.5. Significance Test for the Change Trend of AR in Each Year.

The average CAR in window period for each year (Table 6) was calculated by the announcement date of every M&A event based on the Elman model. The Friedman test results ( $P$  value of significance test is 0.0008) showed significant differences among the AR series in the 14 years, indicating that the M&A short-term performance had a significant downward trend and tended to be more reasonable. The low CAR in 2016–2018 was related to the stock market overall

downturn and the regulatory measures to crack down on the speculation of M&A in the past years, and M&A transactions of listed companies tend to be rational.

5. Conclusions

5.1. Comparison of Algorithms for AR. The stock return series does not have normality, and the assumptions of the traditional regression model cannot be established. In the past research, the algorithm for AR mainly adopted the market adjustment method and the market model method. The former lacked theoretical basis, and the latter was short of statistical basis, affecting the calculation accuracy of AR. To achieve the minimum variance, the market model method

has to take the average value of returns in observation period as the fitting result, and the market adjustment method is an extreme case under this rule, which can be reflected from the empirical result of  $R^2$  0.5346 and 0.2813, respectively. Therefore, the smaller the fitting result, the better the fitting effect, which leads to the predicted normal return underestimated and the CAR overestimated. These two methods have CAR 16.28% and 18.03%, respectively, and significantly greater than other three method's CAR.

In the first four traditional regression models, the unary nonlinear model has the best fitting effect with  $R^2$  0.5459, higher than other three models, because it can deal with nonlinear problems in stock yield time series to some extent, which can also be confirmed by the result that its CAR 14.65% is closest to Elman model' CAR 14.45%.

Besides this, the fitting effect of unary polynomial model is inversely proportional to its prediction ability. With the increase of order, the fitting effect is gradually optimized with  $R^2$  rising from 0.0163 in 4th order to 0.6790 in 10th order, while the higher order unary polynomial model' ability to predict is lost due to its huge volatility because the unary polynomial model cannot deal with the data distribution that stock yield time series fluctuates intensively in narrow numerical range.

The empirical results show that Elman neural network model is capable of solving nonlinear complex problems. It can fit the observation period data as well as predict the AR in window period with  $R^2$  0.9950 and CAR 14.45%, which is significantly different from another 4 traditional regression models. However, Elman neural network model has shortcomings. Only the feedback of hidden layer information is considered in the structure, and the output layer information is not relearned [70]. In addition, although the neural network calculation logic is reasonable and easy to understand, the computer operating process is more like a black box, and few researchers have the ability to analyze the model code to explain why the fitting effect is so good.

The advantages and disadvantages among five calculation models for AR are revealed in Table 7.

**5.2. Short-Term M&A Performance.** The M&A short-term performance in the past 14 years has generally declined, which is related with the effectiveness of China's regulatory measures for M&A hype and speculation, leading to M&A short-term performance tending to be reasonable. The M&A of listed companies can create a 14.45% CAR of considerable short-term performance during the window period, indicating that the stock market is generally in recognition of listed companies' M&A activities and the expectations of the company's value is raised and reflected in stock price. However, in the process of stock price fluctuation, there was hype and speculation behavior on M&A, which was reflected in the fact that the M&A news was transmitted to the stock market at least 2 days before the announcement date, causing the stock price to rise significantly in advance. In addition, the significant change of AR has been prominent for 3 consecutive days after the announcement day. The long-term trading suspension (1302 M&A events were

suspended for 111 days on average) did not digest the centralized or excessive response of the stock market to M&A news.

In addition, the AR is significant for 3 consecutive days after the announcement day, and investors can make use of the public information to obtain excess returns, which proves that China's stock market has not reached the semistrong form of efficiency.

**5.3. Explanation of Chinese Short-Term M&A Performance.** Chinese short-term M&A performance is highly related to the characteristics of China's stock markets. In the process of developing the real economy, China endows M&A great potential and space for value creation, which is reflected either in the economic structural adjustment at the macrolevel or in the industrial transformation-upgrading at the microlevel. The China Securities Regulatory Commission, undertaking the economic management functions as a government department [71], have formulated a series of supportive policies for M&A transactions since 2006, which gives M&A activities the crucial role in capital resource allocation. Besides, different from the western stock markets, which are dominated by institutional investors, China's stock market has 160 million individual investors, accounting for 99.76%, who are more likely to interpret M&A as a good signal and chase the stock, which leads to Chinese listed companies more willing to carry out M&A activities.

**5.4. Enlightenments and Recommendations.** For the majority of researchers, it is recommended to make full use of artificial intelligence method to explore nonlinear problems and compare innovation research with previous research in accordance with statistical principles, which will contribute to improving the rigorism of research modeling and the accuracy of the research conclusion.

## Data Availability

Some or all data, models, or code generated or used during the study are available in a repository or online in accordance with funder data retention policies. All M&A events of listed companies between 2006 and 2019 are from the Wind M&A database, and all stock prices come from CITIC Securities Stock Trading Software.

## Disclosure

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Conflicts of Interest

The authors declare that there are no conflicts of interest.

## References

- [1] Y. He and J. Wang, "The history, current situation and development suggestions of M&A in China," *Coastal Enterprises and Science & Technology*, vol. 1, pp. 41–45, 2011.

- [2] H. Lyu and W. Wang, "Individual financial advisor's reputation concern and M&A performance: evidence from China," *Pacific-Basin Finance Journal*, vol. 60, 2020.
- [3] M. C. Jensen and R. S. Ruback, "The market for corporate control: the scientific evidence," *Journal of Finance Economics*, vol. 11, pp. 5–50, 1983.
- [4] W. Rodermann and M. Gabler, *Strategisches Synergie Management*, Oxford Press, England, UK, 1999.
- [5] M. J. Conyon, S. Girma, S. Thompson, and P. W. Wright, "Do wages rise or fall following merger?," *Oxford Bulletin of Economics and Statistics*, vol. 66, no. 5, pp. 847–862, 2004.
- [6] H. G. Manne, "Mergers and the market for corporate control," *Journal of Political Economy*, vol. 73, no. 2, pp. 110–120, 1965.
- [7] K. Kubo and T. Saito, "The effect of mergers on employment and wages: evidence from Japan," *Journal of the Japanese and International Economies*, vol. 26, no. 2, pp. 263–284, 2012.
- [8] A. Qin Hua, "Theoretical report on corporate mergers and acquisitions," *Corporate Herald*, vol. 6, p. 27, 2012.
- [9] M. Du, A. Boateng, and D. Newton, "The impact of state ownership, formal institutions and resource seeking on acquirers' returns of Chinese M&A," *Review of Quantitative Finance and Accounting*, vol. 47, no. 1, pp. 159–178, 2016.
- [10] A. Noicharoen, *An Empirical Study of Target Firms' Operating Performance after Corporate Acquisitions: Evidence from the Listed Companies in the Stock Exchange of Thailand*, Alliant International University, San, Diego, CA, USA, 2006.
- [11] M. Zhang, *Cross Border M&A Model and Performance of Listed Companies in China*, Jilin University, Changchun, China, 2017.
- [12] X. Zhang, "Does M&A create value?" *Economic Research*, vol. 6, pp. 20–29, 2003.
- [13] P. Madden Gerald, "Potential corporate takeovers and market efficiency: a note," *Journal of Finance*, vol. 36, pp. 1191–1197, 1981.
- [14] P. M. Healy and K. G. Palepu, "Does corporate performance improve after mergers?" *Journal of Financial Economics*, vol. 31, p. 135, 1992.
- [15] A. Agrawal, J. F. Jaffe, and G. N. Mandelker, "The post-merger performance of acquiring firms: a re-examination of an anomaly," *The Journal of Finance*, vol. 47, no. 4, pp. 1605–1621, 1992.
- [16] M. Humphery-Jenner and R. Powell, "Firm size, sovereign governance, and value creation: evidence from the acquirer size effect," *Journal of Corporate Finance*, vol. 26, pp. 57–77, 2014.
- [17] P. Dodd, "Merger proposals, management discretion and stockholder wealth," *Journal of Financial Economics*, vol. 8, pp. 105–137, 1980.
- [18] C. Higson and J. Elliott, "Post-takeover returns: the UK evidence," *Journal of Empirical Finance*, vol. 5, pp. 27–46, 1998.
- [19] S. Hans, *Mergers and Concentration Policy*, Edward Elgar Publishing, Cheltenham, UK, 2006.
- [20] B. N. Jarrel, "Inter firm tender offers and the market for corporate control," *Journal of Business*, vol. 53, pp. 345–376, 1988.
- [21] R. F. Bruner, "Does M&A pay: a survey of evidence for the decision-maker," *Journal of Applied Finance*, vol. 12, no. 1, pp. 48–68, 2002.
- [22] K. C. Yook, "The measurement of post-acquisition performance using EVA," *Quarterly Journal of Business and Economics*, vol. 43, pp. 67–83, 2004.
- [23] M. Uddin, "An analysis of short-run performance of cross-border mergers and acquisitions: evidence from the UK acquiring firms," *Review of Accounting and Finance*, vol. 8, pp. 431–453, 2009.
- [24] S. Li and Y. Chen, "The wealth effect of mergers and acquisitions of listed companies," *Economic Research*, vol. 11, pp. 28–35, 2002.
- [25] X. Liu, X. Huang, and H. Guo, "An empirical study of industrial cycle, M&A types and M&A performance," *Financial Research*, vol. 3, pp. 135–152, 2009.
- [26] J. Deng, Y. Zeng, and J. He, "The roots and consequences of related M&A reorganization," *Journal of Management*, vol. 8, pp. 1238–1246, 2011.
- [27] B. Zhang and Q. Sheng, "Analysis of M&A performance of internet finance enterprises," *Finance and Accounting Monthly*, vol. 29, pp. 46–51, 2016.
- [28] W. Li and He Song, "Research on M&A performance and impact mechanism of listed companies supported by venture capital," *Accounting Research*, vol. 6, pp. 60–66, 2017.
- [29] Z. Zhang and J. Lei, "Is the interest of the company's purchase stakeholder balanced?—analysis of risk premium arbitrage based on the company's purchase incentives," *Economic Research*, vol. 6, pp. 30–37, 2003.
- [30] H. Zhu and Z. Chen, "Research on M&A performance of enterprise technology based on synergistic effect—a case study of listed companies," *Soft Science*, vol. 30, pp. 58–61, 2016.
- [31] X. Chen and T. Zhang, "Market reaction of asset restructuring—an empirical analysis of assets reorganization in shanghai stock market in 1997," *Economic Research*, vol. 9, pp. 47–55, 1999.
- [32] G. Yu and R. Yang, "Theoretical analysis and empirical analysis on enterprises' M&A effect in stock price," *Contemporary Finance and Economics*, vol. 7, pp. 70–74, 2000.
- [33] L. Yu and Y. Liu, "An empirical analysis on M&A performance of China's listed companies," *Contemporary Economics*, vol. 7, pp. 68–74, 2004.
- [34] Y. Chen, W. Chen, and J. Lin, "Research on the reform of control market system, corporate governance, and M&A performance," *Securities Market Herald*, vol. 2, pp. 37–46, 2017.
- [35] J. Deng, Y. Zeng, and J. He, "Associated M&A and reorganization: roots and consequences," *Journal of Management*, vol. 08, pp. 1238–1246, 2011.
- [36] Z. Peng and J. Chen, "A literature review of enterprise M&A performance," *Communication of Finance and Accounting*, vol. 3, p. 83, 2011.
- [37] R. Ball and P. Brown, "An empirical evaluation of accounting income numbers," *Journal of Accounting Research*, vol. 6, pp. 159–178, 1986.
- [38] E. F. Fama, L. Fisher, M. C. Jensen, and R. Roll, "The adjustment of stock prices to new information," *International Economic Review*, vol. 10, pp. 1–21, 1969.
- [39] R. M. Caves, "Takeovers and economic efficiency: foresight vs hindsight," *International Journal of Industrial Organization*, vol. 7, pp. 151–174, 1989.
- [40] J. Fred Weston, L. Mark, J. H. Mitchell, and T. Mulherin, *Restructuring and Corporate Governance*, pp. 152–170, Pearson Education Limited, Harlow, UK, Fourth edition, 1999.
- [41] A. Gregory, "An examination of the long run performance of UK acquiring firms," *Journal of Business Finance and Accounting*, vol. 24, pp. 971–1002, 1997.
- [42] E. Dimson and P. Marsh, "Event study methodologies and the size effect: the case of UK press recommendations," *The Journal of Finance*, vol. 17, pp. 113–142, 1986.

- [43] J. Lakonishok and T. Vermaelen, "Anomalous price behavior around repurchase tender offers," *Journal of Finance*, vol. 45, pp. 455–477, 1990.
- [44] R. G. Ibbotson, "Price Performance of common stock new-issues," *Journal of Financial Economics*, vol. 3, pp. 235–272, 1975.
- [45] W. Cong, *Research on the Impact of Smooth Earnings of Stocks on Stock Abnormal Return*, Lanzhou University, Lanzhou, China, 2014.
- [46] C. W. J. Granger, "Developments in the study of cointegrated economic variables," *Oxford Bulletin of Economics and Statistics*, vol. 48, pp. 213–228, 1986.
- [47] Y. Zhang, "The empirical tests of nonlinear and chaos of Shanghai stock market," *Journal of Industrial Engineering and Engineering Management*, vol. 17, pp. 21–25, 2003.
- [48] C. Su, X. Yan, and X. Liu, "Nonlinear characteristics and fractal dimensional analysis of China's Securities market," *System Engineering Theory and Practice*, vol. 25, no. 5, pp. 68–73, 2005.
- [49] Q. Sun, X. Zhang, and S. Zhang, "A study on the existence of nonlinear cointegration relations," *Journal of Management Sciences*, vol. 3, no. 2, pp. 65–74, 2000.
- [50] R. F. Engle and C. W. J. Granger, "Co-integration and error correction: representation, estimation, and testing," *Econometrica*, vol. 55, no. 2, pp. 251–276, 1987.
- [51] D. Li, S. Wang, D. Li, and X. Wang, "On the theory and method of spatial data mining and knowledge discovery," *Journal of Wuhan University (Information Science Edition)*, vol. 27, no. 3, pp. 221–233, 2002.
- [52] B. Müller and J. Rrinhardt, *Neural Networks: an Introduction*, Springer-Verlag, Berlin, China, 1990.
- [53] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, p. 359, 1989.
- [54] K. Hornik, M. Stinchcombe, and H. White, "Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks," *Neural Networks*, vol. 3, no. 5, p. 551, 1990.
- [55] H. White, "Economic prediction using neural networks: the case of IBM daily stock returns," in *Proceedings of the IEEE International Conference on Neural Networks*, pp. 451–458, San Diego, CA, USA, July 1988.
- [56] M. F. Gonzales, "Modeling market volatilities: the neural network perspective," *European Journal of Finance*, vol. 3, no. 2, pp. 137–157, 1997.
- [57] Ö7 İcan and T. B. Çelik, "Stock market prediction performance of neural networks: a literature review," *International Journal of Economics and Finance*, vol. 9, no. 11, pp. 100–108, 2017.
- [58] S. Weng and X. Lin, "Application of BP network, RBF network, and elman network in stock price forecasting," in *Proceedings of China Intelligent Automation Conference*, Fuzhou, China, May 1999.
- [59] H. Kurt, S. Maxwell, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.
- [60] X. H. Shi, Y. C. Liang, H. P. Lee, W. Z. Lin, X. Xu, and S. P. Lim, "Improved elman networks and applications for controlling ultrasonic motors," *Applied Artificial Intelligence*, vol. 18, no. 7, pp. 603–629, 2004.
- [61] R. Sitte and J. Sitte, "Neural networks approach to the random walk dilemma of financial time series," *Applied Intelligence*, vol. 16, no. 3, pp. 163–171, 2002.
- [62] W. Huang, Y. Nakamori, and S.-Y. Wang, "Forecasting stock market movement direction with support vector machine," *Computers & Operations Research*, vol. 32, no. 10, pp. 2513–2522, 2005.
- [63] J. K. Hyun and S. S. Kyung, "A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets," *Applied Soft Computing*, vol. 7, no. 2, pp. 569–576, 2007.
- [64] X. Han, M. Zhu, and Y. Ji, "Application of OIF elman neural network in stock market composite index prediction," *Journal of Changchun University of Technology*, vol. 27, no. 2, pp. 135–138, 2006.
- [65] Li Ming, X. Han, and L. Wang, "An improved elman neural network and its application in stock market," *Computer Engineering and Applications*, vol. 34, pp. 67–68, 2006.
- [66] C. Lin and D. Zhu, "Research on stock price forecast based on elman neural network," *Computer Applications*, vol. 26, no. 2, pp. 476–477, 2006.
- [67] X. Wang, F. Shi, Yu Lei et al., *30 Case Analysis of MATLAB Neural Network*, pp. 65–67, Beijing Aerospace University Press, Beijing, China, 2013.
- [68] X. Chen and A. Sun, "Validity test of CAPM in China's stock market," *Journal of Peking University*, vol. 4, pp. 28–37, 2000.
- [69] H.-B. ZHU and H. E. Li-Juan, "Research on conditions in the consistency test of normal distribution and others by applying single sample K-S test in SPSS," *Journal of Capital University of Physical Education and Sports*, vol. 4, pp. 466–470, 2009.
- [70] Z. Qi and X. Hou, "BP neural network and its improved algorithm," *Journal of Taiyuan University of Science and Technology*, vol. 26, no. 2, pp. 122–124, 2005.
- [71] Y. Tong and W. Liu, *A Review of the Basic Theories of Government Functions at Home and Abroad*, Jiangxi Normal University, Nanchang, China, 2007.

## Research Article

# The Organizational Structure and Operational Logic of an Urban Smart Governance Information Platform: Discussion on the Background of Urban Governance Transformation in China

Junfang Kang and Xianjun Wang 

*School of Public Policy & Management, China University of Mining and Technology, Xuzhou 221116, Jiangsu, China*

Correspondence should be addressed to Xianjun Wang; wangxianjun@cumt.edu.cn

Received 21 October 2020; Revised 15 November 2020; Accepted 29 November 2020; Published 12 December 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Junfang Kang and Xianjun Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

At present, the Chinese government is trying to resolve various social contradictions, such as people's ever-growing need for a better life and unbalanced and inadequate development. To do so, urban governance practices including holistic governance, decentralized and interconnected governance, multiple participatory governance, and smart governance have been developed in China. Urban smart governance supported by mobile Internet, the Internet of Things, quantum computing, big data, artificial intelligence, and other information technologies has also entered the field of vision of academics and administrators. However, the research and practice on the integration of organizational structure and smart governance technology for urban governance in China are still insufficient. Hence, this paper proposes the design of the organizational structure of an urban smart governance platform and presents the "1 + N" integrated smart information platform and the "power-sharing linkage, one-center and dual-track" platform construction command system. In this system, the functional module of the information platform for smart urban governance is elaborated. This paper contributes to promoting the modernization of China's urban governance capacity, which can enhance social equity and social order, promote social democracy and rule of law, and enhance the efficiency of urban governance.

## 1. Introduction

On April 19, 2016, Chinese President Xi Jinping stated "In the process of promoting the modernization of the national governance system and governance capacity through information technology, we need to coordinate the development of e-government and build an integrated online service platform to perceive social trends, smooth communication channels and assist scientific decision-making" [1]. Since then, the informatization of urban governance has boomed in China. In addition to the government's active promotion, the rapid development of and changes in Chinese cities and the promotion of information technology have accelerated the informatization of urban governance. This paper aims to identify the trends of China's urban governance transformation to design a governance organization structure and to elaborate on the structure's basic operational logic according

to the information platform of urban smart governance. "1 + n" smart city takes "wisdom, wisdom and efficiency" as the construction concept, relying on the core technologies such as artificial intelligence, blockchain, and cloud computing, to build a "1 + n" smart city platform system, with a set of "smart city cloud" platform to strongly support *n* smart city plates, including smart life, government affairs, transportation, education, health, hospital, food safety, port, environmental protection, pension, law, and community.

## 2. Research Background and Literature Review

At present, many countries have begun to build smart cities, mainly in the United States, Sweden, Ireland, Germany, France in Europe, and China, Singapore, Japan, and South Korea in Asia. The construction of smart cities in most countries is in the stage of limited scale and small-scale

exploration. As the world's fourth largest manufacturer of electronic products, South Korea is one of the leading countries in the formulation of international standards for the Internet of Things, cultivating new industries through the construction of smart city. The United States has raised the construction of smart city to the height of national strategy and made key investment and construction in infrastructure and smart grid. Singapore is recognized as the country with the best government service. Information and communication technology promotes economic growth and social progress. The construction of smart city focuses on serving the public [2].

China's smart city construction has just started, and the urban informatization construction is in an important structural transformation period, that is, from the stage of information technology popularization and application to the stage of information resources development and utilization. China is actively making use of the latest technologies such as the Internet of Things and cloud computing to promote the construction of smart city through the strategic deployment of "integration of industrialization and industrialization," "simultaneous development of five modernizations," and "integration of three networks." At present, there are three construction modes of smart city construction in China, which are the construction mode driven by the development of Internet of Things industry, such as Wuxi; the construction mode guided by the construction of information infrastructure, such as Wuhan; and the construction mode with social service and management application as the breakthrough, such as Beijing and Chongqing.

Analyzing the connotation of the word "smart" is the premise of understanding the concept of "urban smart governance." The word "smart" has different origins and meanings. In Chinese, "smart" [3] means "the ability to distinguish right from wrong and to invent something." In English, it refers to "having or showing a high degree of mental ability" or "witty, clever" [4]. These two definitions convey two meanings: the ability to perceive things and the ability to manage things, respectively. Smartness is the key to government reform strategies worldwide [5].

Smart governance can be defined as "a deployment of the creative mix of emerging technologies and innovation in the public sector" [6] that can cope with complex and difficult challenges to promote innovation, sustainability, and competitiveness in society [7]. In recent years, smart governance has attracted increasing attention.

A team of researchers led by Giffinger of the Technical University of Vienna began a comprehensive assessment of the development of smart cities in Europe in 2007 [8]. The level of smart governance in cities was seen as one of the most important aspects of the assessment, which focused on public participation in decision-making, transparency, and public services. Alenezi et al. [9] regarded smart governance as a governance model based on e-government innovation. Rochet et al. [10] believed that smart governance is business processing and information communication based on information and communication technologies (ICT), with the aim to improve service quality. Some researchers [11–14]

agreed that ICT plays an important role in urban governance. Information technology and many disciplinary approaches, for example, have been adopted to help cities improve the use of services and infrastructure and balance the sustainability of social, economic, and environmental impacts. In addition, Ronan [15] noted that smart governance is a government service that is guided by public services and makes full use of various data, information, and advanced ICT to provide smarter and higher-quality services to the public. Nada et al. [16] defined smart urban governance as smart, intelligent urban public service management and public participation. Lv et al. [17] illustrated that ICT has been used by the government to facilitate interdepartmental data sharing and achieve more open and sustainable government services.

With the implementation of the urbanization strategy, increasing attention has been paid to the level of smart governance in China. Related theoretical research and practice have increased simultaneously. According to Cui and Guo [18], urban smart governance takes the whole urban system as the governance object and constructs a smart decision-making mechanism through information technology to make urban public services more standardized, transparent, and efficient through open interaction between the government, markets, and society. Wang [19] reported that urban smart governance covers both management and service, with an emphasis on smart management and control and smart service. Nie [20] emphasized that the government has made full use of big data, mobilized social forces, and formed a pluralistic cogovernance pattern. Gao [21] believed that the use of modern information technology by the government transforms the relationship among the government, citizens, and society and that various departments within the government can realize power sharing, innovative communication, and supervision methods and shape an efficient modern governance model.

In light of the above literature review, this paper defines urban smart governance as the elimination of information islands between government departments, the full use of advanced ICT to collect data in real time, and data sharing and intercommunication between government departments. By processing and analyzing big data through cloud computing and artificial intelligence, we can offer more accurate, smarter, more efficient, and higher-quality government services oriented to public services and ultimately promote harmonious and sustainable urban development.

Academia generally believes that information platforms of urban smart governance are the key to bridging the gap between technology and governance [22, 23]. However, the unreasonable design of existing information platforms for urban governance leads to a gap between data collection and application [24]. Urban smart governance is not the function of a single department but involves many departments. If the government relied on the promotion of one department alone, it would not be able to succeed. However, the disorderly participation of all departments would lead to difficult coordination. Therefore, a unified deployment of the city government is needed [25]. In the management of the Chinese government's data resources, there are two main

problems: the isomorphism of responsibilities and the highly fragmented and segmentary system. The independence and estrangement between the upper and lower levels of government departments, as well as the conflicts and contradictions between municipal departments at the same level, lead to a situation in which horizontal data can be neither aggregated nor shared and vertical data can be shared neither upward nor downward [26].

At present, the research on urban smart governance is still in the stage of theoretical exploration, and there are relatively few studies and practices related to smart governance considering the characteristics of China's urban governance system. Urban smart governance is a systematic project, and it is urgent to integrate information technology and governance systems, improve the existing governance system through the construction of urban smart governance information platforms, and form an organizational structure and governance model of urban smart governance with clear rights and responsibilities.

### 3. The Trend of Urban Governance Transformation in China

In the academic field, "smart governance" is a concept that has been widely discussed and spread in recent years. It is generally discussed in the context of governance capacity and systems. Therefore, the discussion on urban smart governance must be based on the practical problems of urban governance; otherwise, the academic discussion will lose a practical basis. It not only relates to technical problems but also includes the concept, values, and behavioral process of urban governance.

Correspondingly, the current situation and trend of urban governance in application need to be analyzed in advance in order to clarify the real demand for urban governance and further solve problems.

*3.1. Trend One: Holistic Governance.* The purpose of China's "comprehensive urban management" system is to transform urban governance from decentralized to comprehensive governance. The main characteristics of a comprehensive urban management system are embodied in the functions of "centralization" and "linkage." Centralization aims to provide good services, improve management, enhance law enforcement, and centralize the management and punishment power involved in urban management into one functional department in order to integrate urban management and administrative law enforcement and to improve administrative and service efficiency. Linkage refers to the implementation of interdepartmental joint law enforcement for some urban management law enforcement matters that are temporarily difficult to concentrate on in the process of urban management operation, which is embodied in the linkage between similar or related departments. However, the comprehensive urban management system is only a transitional measure in urban governance. To realize the goal of refined management of diversified subjects in a city, government departments must achieve seamless

professional coordination and smooth information transmission. Therefore, the government should embrace the concept of holistic management guided by the needs of citizens, use information technology as the governance means, and choose coordination, integration, and responsibility as the governance mechanisms to carry out organic coordination and integration of governance levels, functions, public-private relationships, and information systems and to address other fragmentation issues [27].

Urban governance system reform based on the concept of holistic governance will ultimately establish an administrative system coordinated by integrating service, management, and law enforcement [28]. Configuration of power among local government departments should become into notice, and relevant rights and responsibilities need to be clarified to establish a mechanism for information communication among various functional departments. In addition, a multidepartmental coordinated urban governance pattern should be adopted to strengthen the evaluation of and feedback on the effectiveness of urban governance in order to ensure that urban governance meets the needs of all parties. With the development of information technology, holistic management is not only characterized by "concentration" and "linkage" but also emphasizes governance by means of the full use of information technology and the concept of serving the public. Moreover, it coordinates the function and accountability mechanisms of information systems, governance classes, and public and private sectors to promote collaboration among public service subjects and provide seamless public services to the public [29]. Therefore, holistic governance will inevitably facilitate the development of participatory governance and smart governance in practice.

*3.2. Trend Two: Decentralized and Interconnected Governance.* Since the beginning of the 21<sup>st</sup> century, the central government of China has gradually transferred its power to lower levels. "New affairs" generated in the process of urban development have also enabled local governments to naturally acquire power. Thus, city governments have acquired the dominant position of urban governance power. The governance of China's city-level government long used management as its main means, and it has long been featured by comprehensive law enforcement. Local urban management and law enforcement departments belong only to institutions under the jurisdiction of local governments, and even if the government sets up branches, they are basically only at the district level. The function of urban governance of streets and communities, especially their position in the governance system, has not received enough attention. A governance system of decentralization with a unified vertical linkage of "city government-district-street-community" has not been established.

With the development of modern information technology, government reform campaigns have been triggered by the data-opening movement in China. The introduction of information technology has changed the organizational system of the government, thus requiring government



management to change from the previous bureaucratic system to a new organizational framework based on networks and technology [30]. Driven by information technology and the rapid development of cities, the change in the organizational structure is profound; it involves pushing the external boundary of the power structure.

Through our long-term study, we found that in recent years, many city governments have defined the scope of urban management authority because Chinese cities have a principal position in the power of urban governance. City governments have clarified the governance rights and responsibilities of cities, districts, streets, and communities, and the decentralized and interconnected governance model is gradually taking shape in Chinese cities. The reform of urban governance in all parts of China has highlighted the difference in the division of functions between decentralized governance and governments at different levels, with the aim to achieve good coordination between municipal governments and governments at various levels within their jurisdictions based on the division of functions (ref). Municipal governments have systematic plans to promote urban governance within the scope of cities. District governments, street departments, and communities, in accordance with their hierarchical differences, exercise their functions and powers in urban governance within their jurisdiction and cooperate with each other to promote the continuous improvement of urban governance.

*3.3. Trend Three: Multiple Participatory Governance.* Urban refined management needs to be coordinated by scientific and effective governance systems and mechanisms. There are three typical models of urban governance [31]: government-centered, market-centered, and user-centered. The purpose of urban smart governance is to break the previous single-center, hierarchical, and inefficient governance mechanism and build a new system of collaborative governance that includes intergovernmental governance, public-private governance, and government-society governance.

With the advancement of China's urbanization development and the formation of the social multiplication pattern, the demands of different group interests are expressed more strongly, which has prompted the adjustment of government decision-making to meet the demands of the public interest and responses. The concept of participatory governance encourages the public to participate in urban governance [32]. As the public is the owner of a city, the development and change occurring in the city are closely related to public life. Encouraging the public to participate in urban governance is the meaning of reforming urban management and improving the level of urban governance in the new era. To attract public participation, the first step is to stimulate the public's willingness to participate, and the second step is to smooth the channels of public participation and innovate the ways of public participation. At the same time, it is necessary to establish and improve the feedback mechanism and establish a long-term mechanism of public participation. An urban smart governance information platform is undoubtedly the best way to meet the above demands.

*3.4. Trend Four: Smart Governance.* As mentioned above, urban smart governance refers to integrating urban governance resources by constantly innovating digital urban management information platforms; making full use of technologies such as the Internet of Things, big data, cloud computing, and mobile Internet to build smart governance platforms; and constructing appropriate systems to improve the service capability of platforms. Smart governance not only emphasizes the management method of digitization but also focuses on expanding the visual and controllable scope of urban governance by means of informatization and digitalization to provide better urban public services, enhance public participation, and realize public value through coordination from management to service. In terms of the governance system, smart governance emphasizes the combination of the government's instrumental rationality and governance's value rationality and pays attention to the transformation of the urban social economy and environment in overall planning to optimize the urban environment and create more suitable living spaces for citizens [33].

The rapid development of information technology in China has affected all aspects of social life and laid the foundation for the practical application of urban smart governance. China is in the information technology era, accompanied by rapid development. Urban governance systems should actively introduce technologies such as big data, cloud computing, and the Internet of Things. With the support of modern information technology, the government can collect and analyze data to grasp the dynamic change process of urban governance for decision-makers in order to better summarize the development and governance rules of urban problems, clarify the follow-up effects of each urban governance measure, and grasp urban development trends. In addition, it is convenient for the government to find problems at the microlevel to improve the wisdom and refinement of urban governance means [34].

*3.5. Disadvantages of Current Urban Governance.* At present, it is difficult for the data mastered by various departments of the city government to be "open as the normal and confidentiality as the exception." Problems such as lack of data, data closure, and repeated construction of smart platforms are common. This includes not only the external factors such as the lag of the overall urban informatization construction, technical barriers, poor organization, and coordination, but also the internal problems such as the lack of power of the department informatization construction, the data as "private property," and the source of rights. For example, the management foundation of information accounts of a large number of key work such as dismantling illegal and scattered pollution enterprises is weak, the data items and data standards are not unified enough, and the task of multidepartment data aggregation and work standardization is a long way to go; some types of data construction are in the bottleneck position, for example, the existing population signaling data and public security work residence permit data need to be integrated and analyzed to reflect the population characteristics, but there is no

relatively formed solution; a large number of urban and rural planning data need to be integrated into space, but the problem of different base map standards needs to be solved first; some departments, including the development and reform, have done a lot of work in the collection of urban operation data and have strong demands for the cross-sectoral application of data and system. However, there is still a large gap between the barrier-free application of data and the linkage of indicators due to the completion of massive data docking.

*3.6. Coordinated Operation Embodies the Advantages of New Smart City.* “Serving enterprises, government, and citizens” is the original intention of smart city. It can be said that “cooperative operation” is the biggest feature of the new smart city, and the “1+n” platform architecture is the biggest logic of the new-type city intelligent governance. Whether it is government affairs, medical treatment, education, environmental protection, law, pension, community and urban construction, and other more subdivided fields, the new urban intelligent governance has been deeply cultivated for a long time. It is worth mentioning that the new type of urban intelligent governance not only focuses on helping to improve the level of urban intelligence, but also makes continuous investment in smart poverty alleviation, contributing to winning the battle of poverty alleviation in rural areas. The new smart city has improved the end-to-end solution matrix including intelligent medical quality control, intelligent disease prediction, intelligent triage guidance, intelligent medical imaging screening, intelligent auxiliary diagnosis and treatment, intelligent ICU critical management, and intelligent follow-up patient education. It has covered 700+ intelligent disease auxiliary diagnosis and treatment models of common diseases, about 800 million patients, and has a complete medical knowledge map including drugs, diseases, prescriptions, and risk factors.

#### **4. The Design of the Information Platform of Urban Smart Governance**

Based on the previous analysis, the current reform and transformation of urban governance in China are embodied in the framework of decentralized and interconnected governance, where overall governance, participatory governance, and intelligent governance resonate at the same frequency. Such a framework is gradually taking shape. The design of the governance organization structure of an information platform of urban smart governance that meets the actual needs of cities should respond to the transformation trend of urban governance.

*4.1. “1+N” Integrated Intelligent Information Platform.* The methodologies of holistic governance and decentralized interconnected governance are adopted to build an information platform for urban smart governance with the basic characteristics of intelligence, informatization, networking, and precision to realize the goal of a multigoverning model of urban governance.

Rónán et al. [15] pointed out that the connotation of smart government governance includes intelligent public services, intelligent social management, and citizen group participation. The “1+N” integrated intelligent information platform aims to build “1” big data resource center and “N” urban governance platforms. This platform covers the content of smart governance and is an open system that expands and reduces its content according to the actual needs of cities (Figure 1).

The platform includes one center, namely, one big data resource center. The big data center is constructed based on cloud computing technology, which combines virtualization, distributed storage, distributed computing, and other technologies with government affairs to ensure the adaptation of advanced technology and government affair applications. Computing, storage, networks, and other basic resources facilitate capacity expansion according to the needs of the application workload. The overall architecture adopts an open design that is compatible with common equipment, mainstream operating systems, virtualization software, and applications in the industry and reduces development, operation, and maintenance costs. The key point of openness is that the function platform can be increased or decreased with changes in the urban governance function. The platform aims to support multiple users and multiple services to ensure that basic resources can be automatically and dynamically scheduled among different applications and users according to their needs. Meanwhile, different businesses can be isolated from each other to ensure the smooth operation of various businesses. The technology and equipment selected for the construction of the big data resource center are advanced, expandable, mature, open, compatible, reliable, safe, and able to control intellectual property. As the infrastructure and supporting platform of urban smart governance, the big data center integrates urban data resources and enhances governance perception. Its most important task is to serve holistic governance.

In the process of constructing this “1N” comprehensive intelligent information platform, in order to better study the distribution of different types of data information in the system, the machine learning and neural network combination model used on this platform will integrate the characteristics of different data categories. The dynamic feature values are attributed to the same data cluster group. When the data feature values of any two groups in the data are not the same, it means that the two types of data feature information are extremely different, and they will be automatically separated into different levels. The data sets are compared with the eigenvalues of the next data. The actual calculation process of the machine learning model and the neural network algorithm can be regarded as the iterative process of repeatedly decreasing the control parameter value (the unique characteristic information of the data in the database) and performing the machine learning algorithm, as shown in Figure 2. Among them,  $c$ ,  $x$ ,  $u$ ,  $y$ , and  $h$  represent different types of data, and  $F$  represents data processing methods.

In the process of automatically classifying data, the neural network algorithm converts the data feature information predicted by the network traffic into the data

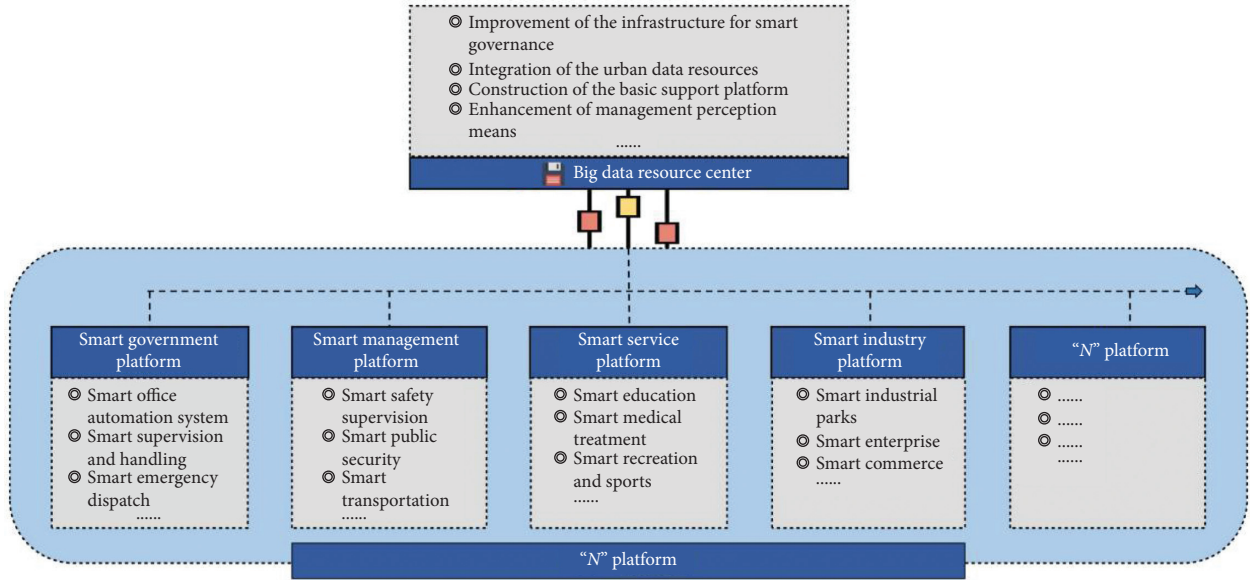


FIGURE 1: "1+N" integrated smart information platform.

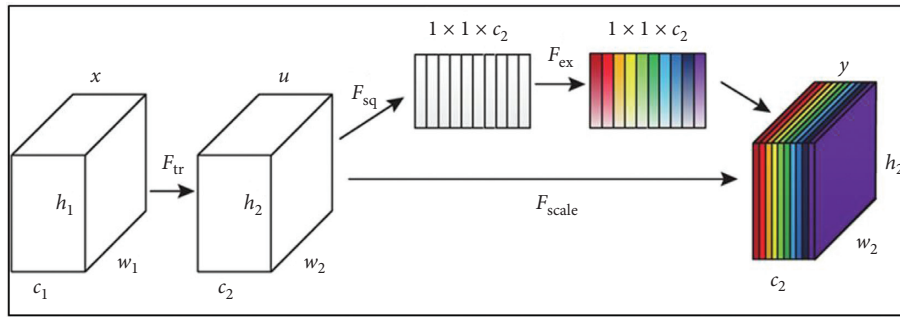


FIGURE 2: Algorithm iteration process in this platform.

information (such as vector group or matrix) that can be recognized by the computer through the feedback link in the automatic analysis process. Realize the adverse effect of output on input, and then achieve automatic high-precision classification based on feedback control. The network traffic prediction model based on the neural network algorithm, under normal circumstances, will automatically analyze and process the data classification that needs to be queried based on multiple feedback links. When we perform predictive analysis on the platform based on the more commonly used state probability formulas in neural network algorithms, the specific formulas are as follows:

$$P_j(T_k) = \frac{\sqrt[k]{KC_k}}{C_{k+1}}, \quad (1)$$

where  $P$  represents the prediction result and  $C$  represents the different types of data analysis. There is also the influence of data  $T_k$  in data eigenvalues on probability  $P_i(T_k)$ .

When the  $T_k$  is very large, the probability of each state is almost equal. At this time, the neural network algorithm began to perform wide-area search, and the  $P_i(T_k)$  difference expanded as the course data predicted by the network traffic decreased.

$$E_i = \frac{K}{T_K}, \quad (2)$$

where  $E$  represents different dimensional data and  $K$  represents different data selection rules at this time:

$$\sigma_\theta(x) = \sum_{i=1}^n P_i(T_\theta). \quad (3)$$

The validity and error of the problems involved in this study are related to the set error confidence  $h$ , and the correlation with the initial solution is also relatively large. In order to only examine the numerical method itself, usually only the stability of the numerical method used to solve the model equation is tested. The model equation is given by

$$\begin{cases} y' = \lambda y, \\ y(a) = y_0. \end{cases} \quad (4)$$

Among them,  $\lambda$  is a complex number, this equation is also called the test equation, and its true solution is as follows:

$$y(x) = y_0 e^{\lambda(x-a)}. \quad (5)$$

When using the random probability model to solve different practical problems, the result is

$$(1 - h\lambda\beta_k)y_{n+k} = \sum_{i=0}^{k-1} (\alpha_i + h\lambda\beta_i)y_{n+i}. \quad (6)$$

Let the solution be

$$y_n = r^n. \quad (7)$$

Then, there is

$$(1 - h\lambda\beta_k)r^{n+k} = \sum_{i=0}^{k-1} (\alpha_i + h\lambda\beta_i)r^{n+i}. \quad (8)$$

Its equivalent form is given by

$$(1 - h\lambda\beta_k)r^k = \sum_{i=0}^{k-1} (\alpha_i + h\lambda\beta_i)r^i. \quad (9)$$

We call the above formula the feature confidence solution formula based on probability random variables and their numerical features. Remember

$$\pi(r; h\lambda) = (1 - h\lambda\beta_k)r^k - \sum_{i=0}^{k-1} (\alpha_i + h\lambda\beta_i)r^i. \quad (10)$$

And take the above formula as the limit characteristic error degree in solving the probability distribution model.

Therefore, when the optimal solution of the network traffic prediction system needs to be solved, the output image of the function when the parameter  $\theta$  is 1 and 2 is shown in Figure 3 (using different ST functions for simulation), where the horizontal axis  $N$  represents different sizes in the dataset and the vertical axis  $T$  represents the accuracy calculation operator.

The platform includes  $N$  platforms, namely, all platforms for building smart government affairs platforms, such as the Smart Government Platform, the Smart Management Platform, the Smart Service Platform, and other platforms required by holistic governance. For example, (1) the Smart Government Platform takes handling documents, meetings, and affairs as the core, and it meets the needs of the Party and government for coordinating government affairs. Paperless office and mobile approval support and data visualization make decision-making more scientific. (2) Through the Smart Management Platform, the government adopts the combination of the Cell Grid Management Method and the Urban Component and Event Management Method and the tools of real-time information collection and transmission to reconstruct the urban management process in order to realize accurate, agile, efficient, and multidimensional urban management. (3) The Smart Service Platform, adhering to the concept of Internet + government services, aims to improve the level of social governance refinement and comprehensively promote networked and highly informationized public management and services from a higher starting point. This platform innovates the government management model, clarifies responsibilities, and integrates resources to achieve “zero distance in social services, full coverage in social governance, and prompt responses to residents” demands. It

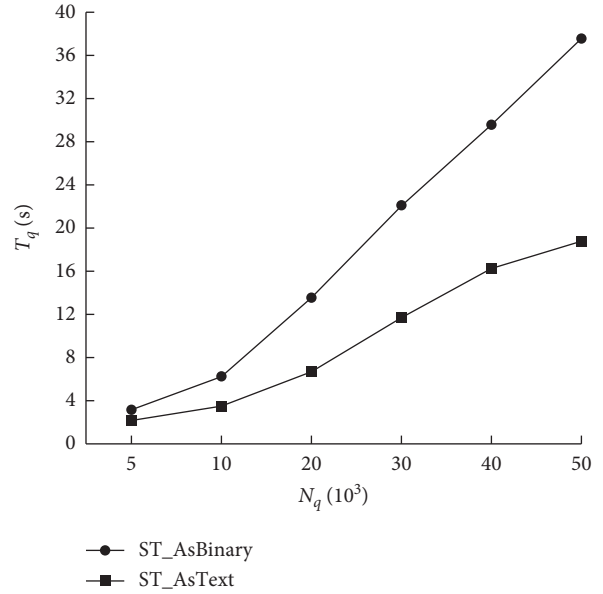


FIGURE 3: Function output image during simulation.

realizes the transformation of urban governance from prevention and control to humanized and service-oriented governance and from extensive governance to refined governance.

**4.2. “Power-Sharing Linkage, One-Center and Dual-Track” Platform Construction Command System.** As shown in Figure 4, power-sharing linkage refers to the distribution of jurisdiction according to the functions of different levels and departments. The smart governance system can automatically allocate and fully link public affairs notifications and governance data between different levels and functional departments within the scope of law. “One center” refers to the “Urban Smart Control Service Command Center,” and “dual track” refers to two tracks. The first track is from the Municipal Government Command Center to the Municipal Departments Command Center to the District Departments Command Center to the Street Departments Command Center and finally to the Community Service Center, which serves as the control service command information channel in ordinary times. The other track is from the Municipal Government Command Center to the Street Command Center and finally to the Community Service Center, which is mainly used as an emergency joint command information channel and auxiliary information channel in ordinary times. Accordingly, a “one-center and dual-track” operation carrier with high efficiency is formed to ensure a city’s control and command linkage. The vertical hierarchy and horizontal organizations are closely integrated and integrated through information technology [35].

## 5. Operation Logic of Urban Smart Governance Information Platform

**5.1. Functional Module of Information Platform for Urban Smart Governance.** Urban smart governance is the reform and innovation of traditional government governance under

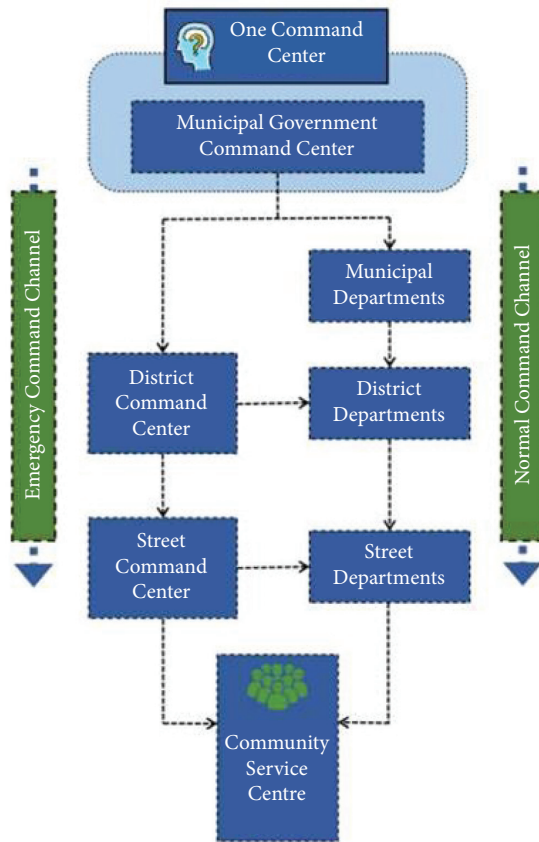


FIGURE 4: “Power-sharing linkage, one-center and dual-track” platform construction command system.

the framework of smart city construction. Gil-Garcia et al. [6] believed that with the help of modern information technology development, smart governance has made creative breakthroughs in government strategy and investment to make governance sustainable and flexible. On this basis, the Smart Information Platform of Urban Governance aims to bridge the islands of information formed by the decentralized storage and management of departments and to form a computer room, a server cluster, a network, and a data center to realize big data collection, access, integration, management, application, and release. Based on the data center, the smart information platform will command and coordinate administrative affairs and urban operation monitoring (Table 1).

## 5.2. Operational Logic of Urban Smart Governance Information Platform

### 5.2.1. Data Integration

(1) *Data Integration.* Driven by the rapid development of information technology, big data has been integrated into people’s lives; it has not only profoundly changed the political-ecological environment of cities but also promoted the transformation of democratic politics from passive participation to active participation. Every governance body is the source of data, and fairness for actors is based on the deep

integration of networks and data. In view of democracy, Decker [36] regarded big data as a “disruptive innovation.” Democratization of data has been quietly taking place, and government construction will develop towards efficiency, innovation, and transparency.

As shown in Figure 5, the platform integrates urban infrastructure, resources, the environment, municipal management, and commerce into a holistic system through the analysis and sharing system and the application of intelligent delivery, communication, Internet, and data processing technology. In addition, it forms a data center through the Internet, cloud computing, data discovery and analysis, and other technical means and constructs subsystems, such as the urban population, transportation, energy, trade, telecommunications, and environmental resources, to perceive, transmit, process, and share data intelligently in the urban system. Urban governance can realize the real-time exchange of data from the system according to different permissions; this enables the quick coordination and arrangement of resources to make governance decisions that allow a harmonious and sustainable city and enable an effective response to unexpected urban security issues.

(2) *Data Collection.* Deutsch [37] explained that information communication is as important to the government as nerves are to the human body. The collection and processing of social information and data assist the government in making scientific and efficient decisions, and the exchange of information inside and outside the system promotes the sustainable operation of government work. Based on the current situation of China’s urban governance transformation, it is recommended that two halls and two centers be adopted as the concrete means of smart governance. It will serve administrative examination and approval of items to achieve the objectives of one window of acceptance, interconnection, and information sharing and to carry out two goals (100% online declaration and 100% online approval). Similarly, the platform is used to reform the administrative system (Figure 6).

All-natural persons, legal persons, and other entities have unique identifiers. That is, legal persons and other organizations have a unified national ID code so that government affairs can be handled with one code, which is conducive to improving the management level of public administration, reducing transaction costs, and improving social work efficiency (Figure 7).

Paperless offices, electronic declarations, and approval are implemented in the process of providing administrative services. For example, when a legal person declares matters, government staff actively search and check the certification materials through the database of the government service exchange platform to avoid trouble for the applicant and to prevent the need for a large amount of paper-version materials and certificates, as shown in Figure 8. The further development of information technology and urban management models is expected to further expand the coverage and efficiency of online services. “More data and less public running” will no longer be a slogan but will be gradually realized in more fields in the future.

TABLE 1: The functional module of information platform for smart urban governance.

| Data integration  | Public affairs allocation processing   | Urban operation monitoring   |
|---|--|--|
| (i) Data acquisition<br>(ii) Real-time data input<br>(iii) Data integration and management<br>(iv) Data applications and services | (i) Network management and unified distribution and disposal center<br>(ii) Three steps: first time to find, first time to deal with, first time to solve<br>(iii) Two separation: patrol separation and handling separation<br>(iv) Five systems: comprehensive inspection, diversified participation, intelligent governance, scientific law enforcement, good credit construction | (i) Government affairs operation monitoring<br>(ii) Urban safety operation monitoring<br>(iii) Economic operation monitoring |

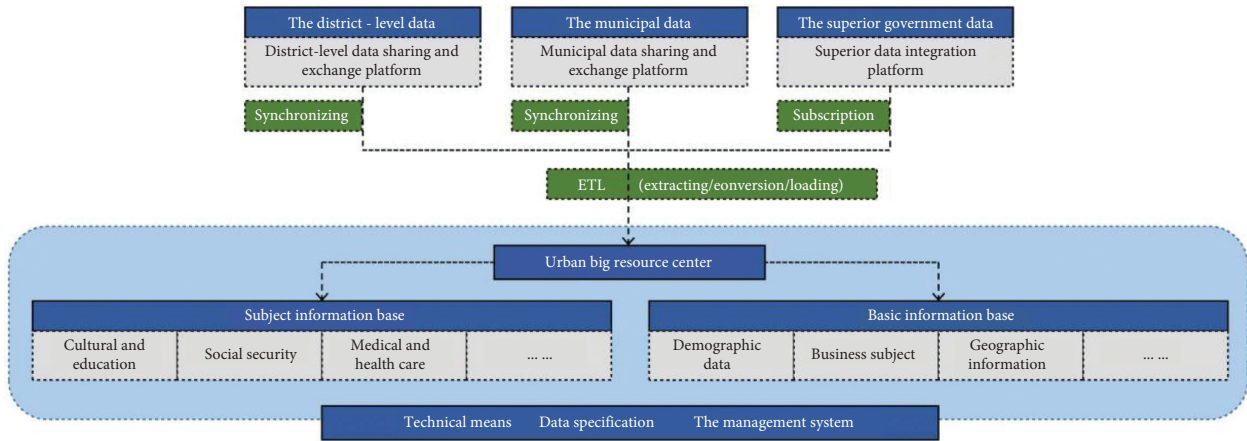


FIGURE 5: Data integration and application logic of the urban smart governance information platform.

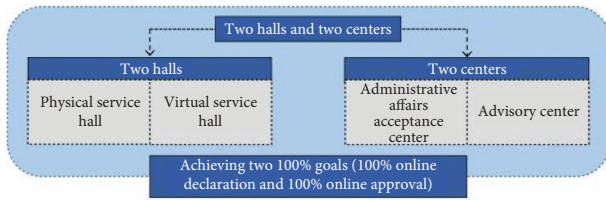


FIGURE 6: Transitional government management channels.

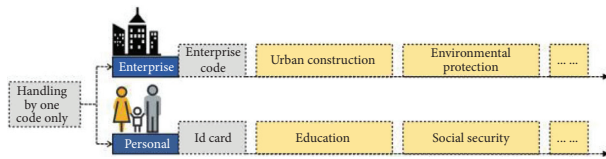


FIGURE 7: Handling by one code only.

5.2.2. Allocation Processing of Public Affairs

(1) *Basic Framework of Distribution Processing Operation.* Although unprecedented social changes have effectively promoted the progress of urban development, they have also created contradictions and challenges at the critical stage of city transitions. To solve different problems efficiently, unified scheduling and distribution processing are needed. The Smart Allocation Processing Platform, based on the principle of separation of inspection and treatment, uses GIS, remote sensing, apps, and other information technologies to realize smart governance for administrative affairs. As shown in Figure 9, the closed-loop processing

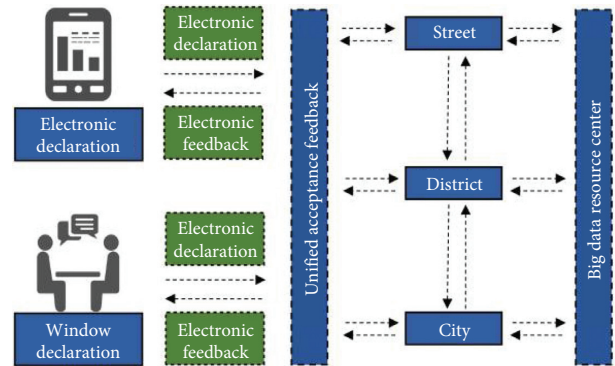


FIGURE 8: E-government service framework and operational logic.

mechanism of “event collection, allocation, disposal, verification, evaluation, and settlement” is formed to establish the information support capability for first-time discovery, first-time disposal, and first-time solutions for urban governance events to promote the refined and scientific development of social governance. The whole operation design combines service classification, spatial information service chain, workflow, and other technologies to ensure clear business logic and convenient operation of transaction processing according to a specific business logic.

(2) *Support System for Allocation Processing Operation.* Relying on the Smart Information Platform of Urban Governance, urban governance areas are divided into community cell grids according to certain standards to form grid governance. Through the inspection and monitoring of

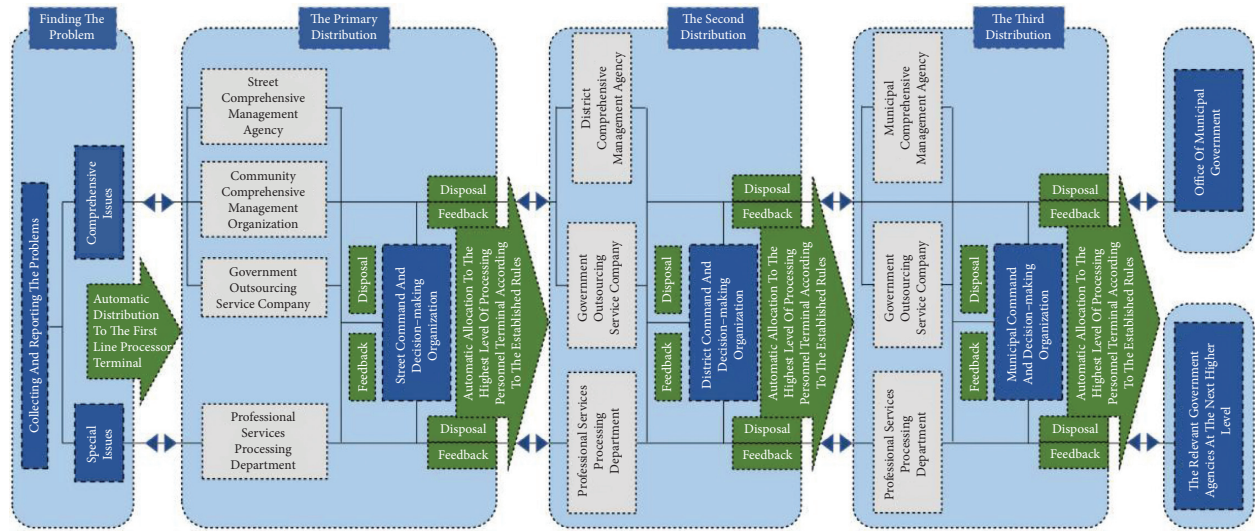


FIGURE 9: Smart distribution processing operational framework.

the community cell grid, unified scheduling, allocation, and disposal, the form of separation of supervision and disposal is established.

Comprehensive inspection forms a comprehensive and professional patrol team at the street level. Through a comprehensive patrol in which one person is in charge of one space and a professional patrol in which one person is responsible for a large patrol network, the whole coverage patrol is formed to identify problems for the first time. This can achieve the target of proactive discovery and timely treatment to improve governance efficiency and solve urban problems in the germination stage. Diversified participation means actively mobilizing the wide participation of society, such as NPC Members, Party Deputies, CPPCC Members, retired cadres, volunteers, and residents' representatives, to enrich the forces for large-scale inspection and renovation. Intelligent governance, the "brain" of urban governance, aims to strengthen the operation and application of digital management means such as data integration, operation monitoring, distribution processing, and collaborative command in the grid management system to integrate all relevant data on people, events, and objects into the system. In addition, it will strengthen commands with big data technology to achieve instantaneous distribution and real-time processing. Scientific law enforcement, forming strong synergy through self-inspection by enterprises, government supervision, and law enforcement investigations to focus on solving problems at the first line, will continue to improve and reform the law enforcement mechanism. Good credit construction, the foundation of management and control, will strengthen the data application of grid management and integrate the credit records of enterprises and individuals into a united credit database and achieve the first correlation in the application and approval of administrative matters. In

this way, each responsible subject is forced to fulfill its own responsibilities (Figure 10).

**5.2.3. Urban Operation Monitoring.** Due to the influence of multiple factors of physical society and virtual society, cities in the information age are faced with increasingly complex unexpected events, and city operation and management are also facing great challenges. As governments attach increasing importance to data governance, they should consider how to change decentralized data management into comprehensive data management using a unified structure, organization, and process. Accordingly, the management and service mechanism using data for decision-making, management, and services will be formed. Urban operation monitoring and early warning play an important role. The application of the Smart Information Platform of Urban Governance in urban operation monitoring will directly or indirectly affect the processes and results of emergency management. The concept of deep integration of technologies such as the Internet of Things, big data, cloud computing, and spatial geographic information technology with business domains of comprehensive emergency management has been accepted and recognized by many municipal governments and academics in China. As shown in Figure 11, City Operational Visualization is the exhibition of real-time situational awareness, collection, sharing, and information about multidimensional interagency collaboration for emergencies and responses to events. This will connect relevant departments, agencies, and command centers at all levels of the emergency response to realize the communication and sharing of real-time situation information in order to ensure joint responses and coordinated emergency responses and to

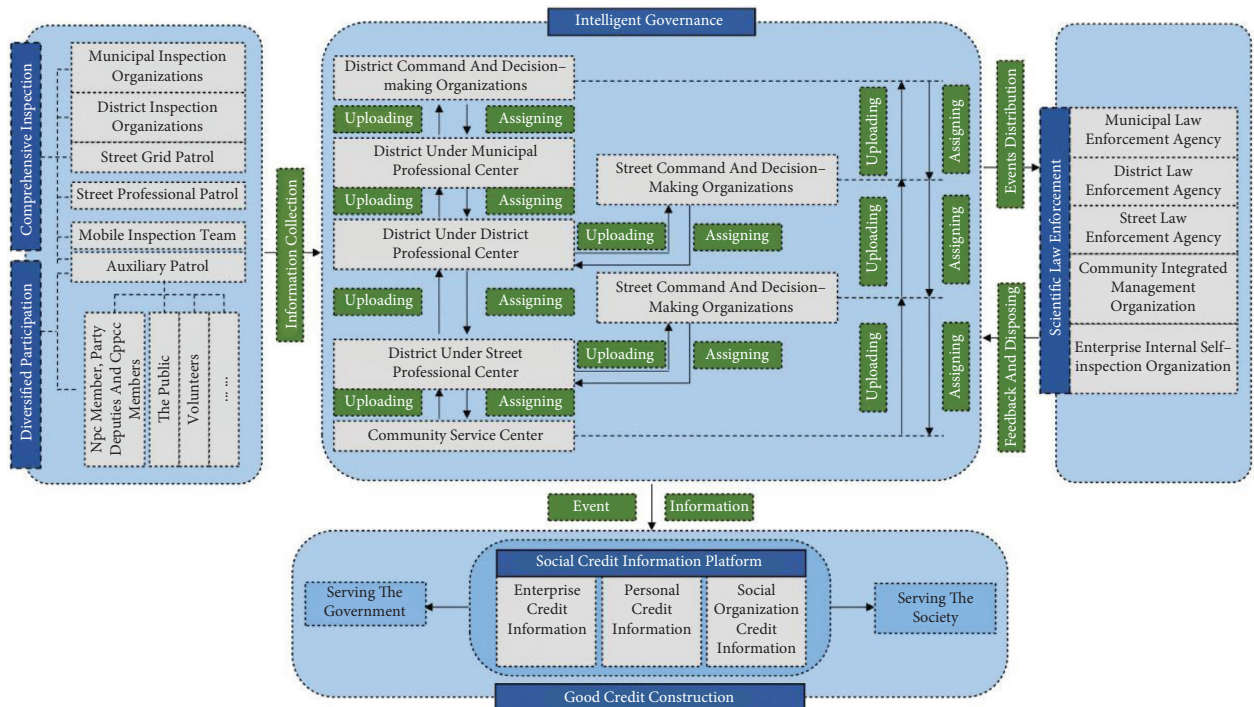


FIGURE 10: Smart distribution processing operational support system.

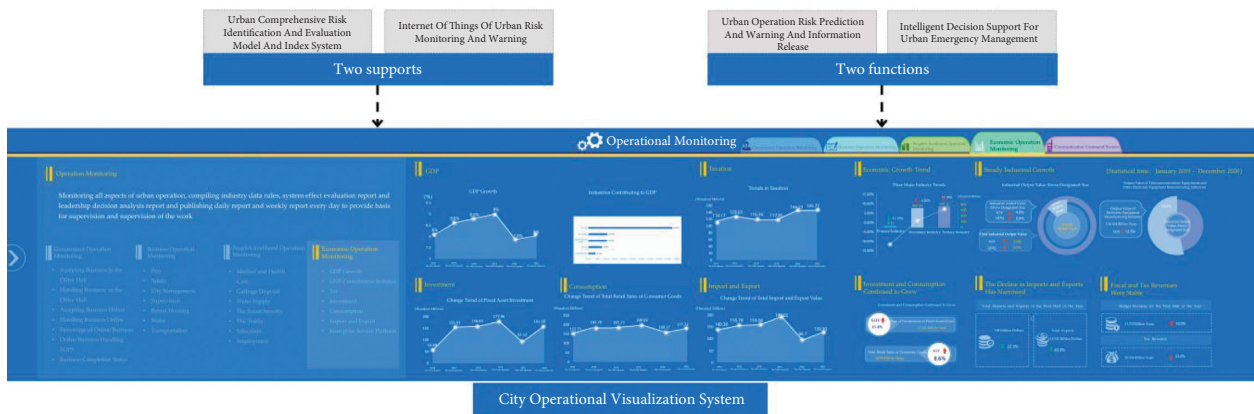


FIGURE 11: City Operational Visualization system.

realize the sharing of emergency resource information and visual command.

Two major supports are as follows:

**Urban Comprehensive Risk Identification and Evaluation**

The risk evaluation model and index system of government affairs, city safety, the economy, and other fields of operation will be established to realize the risk evaluation and analysis of a single field. Hence, machine learning and other technologies are used to continuously learn and train real-time data to realize the automatic optimization of risk analysis models and

intelligent risk prediction. Finally, risk prevention and control measures will be generated automatically.

**Internet of Things of Urban Risk Monitoring and Warning**

The monitoring and early warning system covering all levels, services, and processes in urban governance through the Internet of Things will be constructed. The most significant goal is to strengthen the monitoring areas closely bound to people’s lives, such as urban lifelines, traffic, the environment, flood prevention, and fire control and develop comprehensive application systems for monitoring urban operational signs, risks and hidden dangers, risk assessment,



prediction, and early warnings in order to provide application support for various subjects of urban governance.

Two functions are as follows:

#### Urban Operation Risk Prediction and Warning and Information Release

Artificial intelligence and other information technology are used to achieve higher government efficiency and the scientific analysis of disaster development trends, accident impacts and consequences, economic operations, and so on. First, the early-warning information from various departments will be integrated and strengthen the research and application of precise early-warning release technology to realize the targeted release of urban operation. Second, different responses to early warning persons and the public should be used, and innovation should be developed in the mechanism of early warning information release. The channel of early warning information release should be unblocked, and the effectiveness of early warning information transmission should be improved.

#### Intelligent Decision Support for Urban Emergency Management

According to the types and evolution trends of urban operation emergencies, this paper conducts in-depth mining and analysis of relevant data under multisource data fusion technology for event information and relevant cases to intelligently provide decision-making support for emergency event disposal.

**5.3. Operational Performance Evaluation Mechanism.** Evaluating the process and results of urban governance based on the Smart Information Platform for Urban Governance is important for promoting the modernization of urban governance in the new era. This platform can quickly process acquired data and realize the sharing and linkage of real-time data. It provides the possibility for relevant subjects of urban governance to supervise, evaluate, provide timely feedback, and adjust the content of urban governance. Moreover, the efficient, scientific, and credible modern system of urban governance must rely on the establishment of a performance evaluation mechanism.

On the one hand, the establishment of a sound performance evaluation mechanism can give full play to the external evaluation and supervision role of society and the public in urban governance to adjust governance decisions and to improve government credibility. At the same time, it can improve the sense of participation of social organizations and the public as the main body of urban governance and stimulate the enthusiasm of social organizations and the public [38].

On the other hand, a perfect performance evaluation mechanism can improve the scientific level of urban governance and guarantee the effectiveness of the platform. Due to the real-time and rapid characteristics of information means, policy feedback is timely and convenient for the timely adjustment of urban governance policies and means to make decision-making more scientific and accurate. At

the same time, the serious consequences of decision-making mistakes need to be reduced or prevented.

Therefore, in the era of smart urban governance, it is necessary to build a scientific and perfect performance evaluation mechanism and to build a performance evaluation system during the construction of the Smart Information Platform for Urban Governance to ensure the smooth progress of modernization of urban governance and make urban governance more scientific, more refined, and smarter. Meanwhile, the data resources of various intelligent systems are integrated to analyze the occurrence characteristics, trends, and changes in events. Various kinds of data will be adopted to perform a comprehensive evaluation of the work efficiency of inspection and rectification, to supervise abnormal situations in specific event handling processes, and to provide reference data for the evaluation and evaluation of inspectors.

**5.4. Experimental Results and Analysis.** Table 2 and Figure 12 show the results of 6 sets of data during the work process of the platform's database under the network traffic prediction. According to Table 2 and Figures 12–15, the horizontal axis  $N$  represents different flow data lengths, and the vertical axis  $R$  represents the calculated complex rate.

We can know that in the 5 sets of data, due to the different processing speeds of different types of data, the method based on the combined model of machine learning and neural network is better in terms of accuracy. In terms of computational complexity, we propose an optimization algorithm based on neural network. The complexity of the traffic prediction system is lower, and the difficulty of condition tracking is also lower.

It can be seen from the results that the accuracy of the model proposed in this paper is 0.984, 0.966, 0.967, and 0.932 in terms of data analysis and information extraction. Compared with 0.875, the accuracy of this model can be greatly improved.

On the other hand, in the process of 5 sets of experiments, it can be found that with the different number of experiments, the error of the experiment also shows a regular change, that is, the more the number of experiments, the smaller the prediction error.

## 6. Research and Analysis

### 6.1. Suggestions on Application and Construction of Urban Intelligent Management Platform in China

**6.1.1. Realizing Platform Building and Data Sharing from Top to Bottom.** It is the integration of many data chains within the city that are really concerned about the data links between different departments. Therefore, the construction of the whole platform must be led by the city managers, from top to bottom to strengthen department cooperation, government enterprise cooperation, and government people interaction. Based on the connotation of urban operation and development and urban management, the complete logic chain from macrocontrol to microgovernance is sorted out, and various complex problems involved in urban

TABLE 2: 5 sets of experimental data results.

| Enterprise   | Data 1 | Data 2 | Predicting fit (%) |
|--------------|--------|--------|--------------------|
| Enterprise 1 | 93     | 97     | 87.5               |
| Enterprise 2 | 94     | 94     | 93.2               |
| Enterprise 3 | 89     | 83     | 92.1               |
| Enterprise 4 | 93     | 98     | 96.7               |
| Enterprise 5 | 92     | 97     | 96.6               |

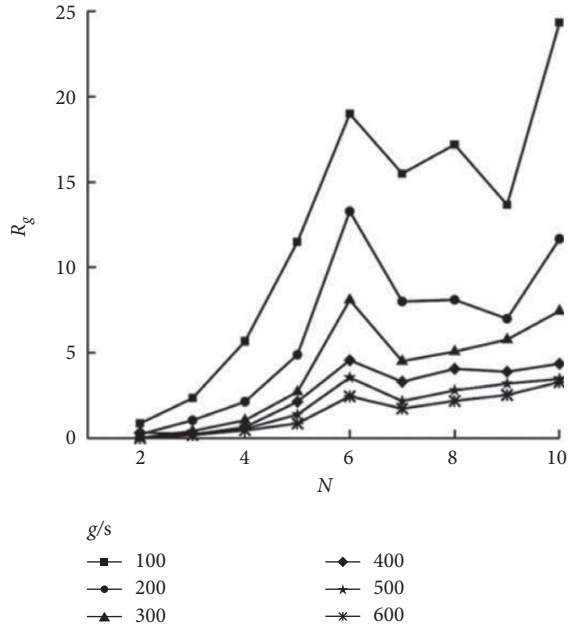


FIGURE 12: Image output of 5 sets of experimental data results A.

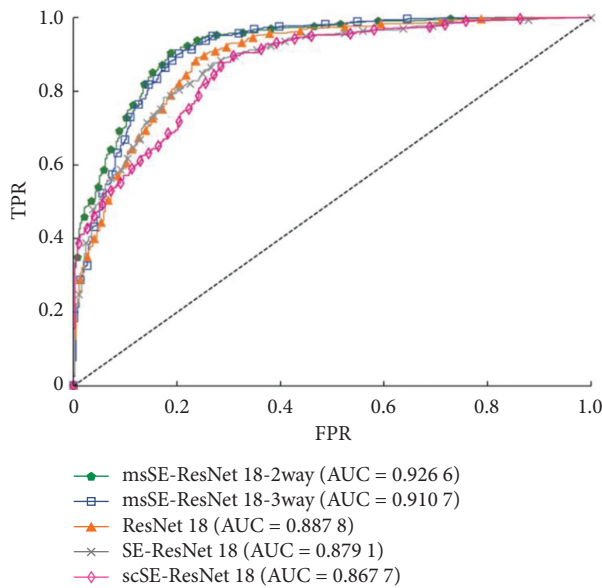


FIGURE 13: Image output of 5 sets of experimental data results B.

management are collected to realize intelligent perception and event response through the platform. The construction of the platform is the work of coconstruction and sharing of multidepartments and social institutions. On the basis of

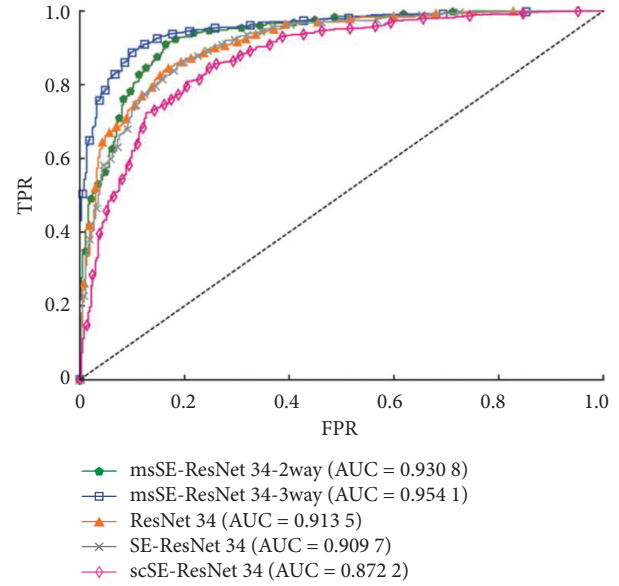


FIGURE 14: Image output of 5 sets of experimental data results C.

“data fusion,” the decision-making of each department needs to comprehensively consider the data and work of relevant departments, and the multiobjective comprehensive decision-making across departments can be realized through data sharing.

**6.1.2. Effectively Improving Data Quality through Data Governance.** The value of data governance lies in standardizing information transmission between systems, improving data quality, and improving the level of data sharing and sharing between business and systems. Scientific data quality rules should be established, and data quality management should be run through all aspects of platform construction and use to meet the needs of business operation and management analysis. Meanwhile, data redundancy and infrastructure investment in storage, hardware, operation, and maintenance should be reduced to improve application system performance and reduce the burden of later maintenance. For some sensitive data, the data items should also be divided into specific sensitive levels, and according to this level, the data items can be properly protected in the whole life cycle, and the protection of residents’ privacy data and key urban infrastructure should be strengthened.

**6.1.3. Designing Scientific Urban Operation Monitoring Index System.** The construction of urban operation monitoring index system should follow the principles of importance, guidance, and systematicness. The selection of indicators should be in line with the urban development strategy and social governance objectives and closely related to the scope of responsibility of urban managers; the construction of the index system must be based on the accurate understanding and linkage analysis of urban governance system, department business, social institutions, and related disciplines, so as to achieve effective coordination among

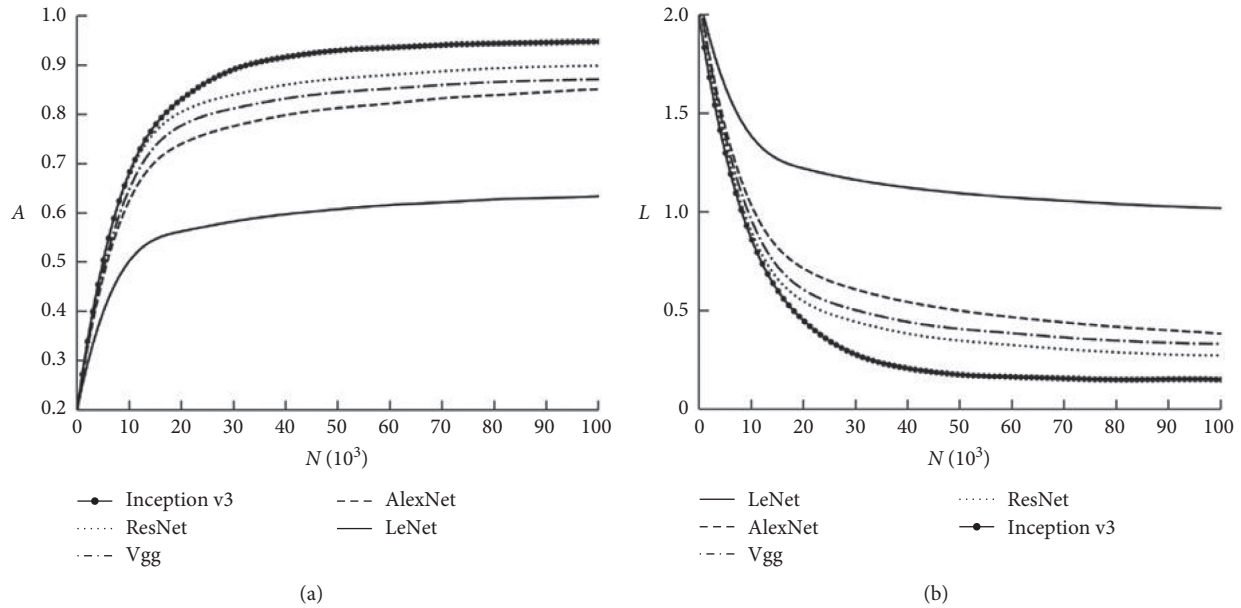


FIGURE 15: Image output of 5 sets of experimental data results  $D$ .

departments and minimize the negative effects and external costs of work.

#### 6.1.4. Building Clear and Practical Application Scenarios.

The application scenarios of urban intelligent management platform should be able to quickly show the city operation status and display the data change trend through the “one plug to the bottom” data granularity and space-time coordinates, so as to effectively support urban management decision-making. The real-time monitoring of abnormal points should be clear at a glance, and the analysis function should be provided according to the needs of users to help them accurately determine the causes of abnormal information. The simulation of urban development scenarios is also one of the government’s key needs for platform capability. Supported by the complex background model, it simulates the development of the city and society, providing the basis for evaluation and decision-making. On the premise of meeting the content, the application scenario design should be clear and concise, the operation function should be simple and practical, and it can provide appropriate personalized customization according to the needs of users and can be used anytime and anywhere by portable devices.

**6.2. Research Contribution.** Technological change and innovation promote the full integration of ICT into the system of political power and urban governance. The superposition and combination of technological elements and other elements are a strong driving force of reform, which has become the catalyst for the development and progress of modern society in China. Wherever it goes, the original power structure system will be deeply changed and form a unique logic of technological governance and development.

Specific technologies and specific governance goals and trends are closely linked to promote the mutual service and promotion of urban governance and information technology in a new and radical way [39]. In this paper, emerging Internet technology and urban governance are integrated for achieving smart urban governance, reengineering business processes, and transforming institutional advantages into governance efficiency.

**6.2.1. Improving Social Equity and Social Order.** The core concept of smart governance is people-oriented, and data-oriented instrumental rationality coexists with humanistic care. The transformation of urban governance from extensive management to precision governance will be promoted by this platform to improve the accurate supply capacity of various public services. Relying on big data acquisition and mining technologies, this platform can seize the demand characteristics of the public, enterprises, and other organizations and improve the matching, precision, and validity of governance to satisfy increasingly diversified and personalized demands and realize the public service of active response to the supply side. In no way does this paper propose conducting urban governance through methods other than science. The research of this paper is based on the existing social system, organizational structure, and urban problems. The famous scholar Hu [40] once pointed out sternly, “We do not deserve to reject science under the background of the chaotic altar monastery, underdeveloped communications, underdeveloped industry.” There is no reason to reject science and technology in China in the present study and application, and even in urban governance, there should not be a one-sided emphasis on the human elements. Instead, we should construct the contemporary value order by means of technology, paying attention to the survival status of all social strata so that the

public can truly share the fruits of development and the problem of social equity can be solved.

**6.2.2. Promoting Social Democracy and the Rule of Law.** It is necessary to establish the operation rules of government data, rebuild the operation rules of government affairs, and reform government institutions and before applying the platform to meet the system and business process requirements of smart urban governance. In this way, the modern urban governance mechanism of adopting data to make decisions, manage public affairs, and innovate will be constructed to adapt to the demand of smart urban governance. Through the integration of business, technology, and data, the government will reshape the government process, organizational structure, functional positioning, and responsibilities, break down organizational and information barriers, and improve the government's governance capacity and public service quality [41]. The platform will work with the new system to promote the participation of all citizens in urban governance to objectively promote the application of democracy. The extensive participation of the public in governance can pool the wisdom of the public to the greatest extent in order to improve the scientific nature of urban governance and establish good governance in the actual urban governance practice that conforms to the fundamental interests of the public.

**6.2.3. Building a Chinese Sample of City Governance.** It is indispensable to improve the efficiency of urban governance in the information age by ICT. The application of the urban smart governance information platform will be conducive to improving the efficiency of government operations, promoting economic development, social governance, and cultural development, and enabling all of society to run like clockwork. The application of the platform, which conforms to the transformation trend of urban governance in China, will set a governance example for Chinese cities.

**6.3. Discussion.** The smart governance concept further promotes the city governance path of innovation. The urban smart governance information platform, an effective means of implementing intelligent management, will constantly improve the management network at the city, district, street, and community levels in practice to form a unified intelligent information platform for the governance of the whole city. Platform governance includes matters in various areas of urban social life and information on people, places, things, events, and organizations within the jurisdiction into the database management. The establishment of a long-term mechanism for data updating and the realization of dynamic monitoring and management should be established to discover and process problems in a timely manner; this enables effectively realizing the government's public management and services to social units. However, it needs to be emphasized that the formation of any urban order cannot be achieved without the participation of people. It is difficult for the government to form a flexible order by means of strong

administrative regulation or wishful planning. Urban governance transformation must be built and operated under the guidance of a people-oriented concept, gathering diverse subjects and various social forces and jointly promoting the modernization of the urban governance capacity.

## Data Availability

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also form part of an ongoing study.

## Conflicts of Interest

The authors declare no conflicts of interest.

## References

- [1] J. Xi, "Speech at the symposium on network security and information technology," 2016, [http://www.cac.gov.cn/2016-04/25/c\\_1118731366.htm](http://www.cac.gov.cn/2016-04/25/c_1118731366.htm).
- [2] China Internet Network Information Center, "The 43<sup>rd</sup> statistical report on internet development in China," 2019, [http://www.cac.gov.cn/201902/28/c\\_1124175677.htm](http://www.cac.gov.cn/201902/28/c_1124175677.htm).
- [3] Online Chinese Dictionary, <http://xh.5156edu.com/html5/z77m56j16220.htm>.
- [4] Webster's new world dictionary, [https://www.cambridge.org/core/search?q=smart&csrf=92UDTC9j-rqf-Mtn8flbHPI\\_JgBQd3IYYVcU](https://www.cambridge.org/core/search?q=smart&csrf=92UDTC9j-rqf-Mtn8flbHPI_JgBQd3IYYVcU).
- [5] S.-J. Eom, N. Choi, and W. Sung, "The use of smart work in government: empirical analysis of Korean experiences," *Government Information Quarterly*, vol. 33, no. 3, pp. 562–571, 2016.
- [6] J. R. Gil-Garcia, N. Helbig, and A. Ojo, "Being smart: emerging technologies and innovation in the public sector," *Government Information Quarterly*, vol. 31, pp. 1–18, 2014.
- [7] H. J. Scholl and M. C. Scholl, "Smart governance: a roadmap for research and practice," in *Proceedings of the iConference*, pp. 163–176, Berlin, Germany, 2014.
- [8] European smart cities, <http://www.smart-cities.eu>.
- [9] H. Alenezi, A. Tarhini, and S. K. Sharma, "Development of quantitative model to investigate the strategic relationship between information quality and e-government benefits," *Transforming Government: People, Process and Policy*, vol. 9, no. 3, pp. 324–351, 2015.
- [10] C. Rochet, J. D. Pinzón, and J. David, "Urban lifecycle management: a research program for smart government of smart cities," *Revista de Gestão e Secretariado*, vol. 7, no. 2, pp. 1–20, 2016.
- [11] D. Pérez-González and R. Díaz-Díaz, "Public services provided with ICT in the smart city environment: the case of Spanish cities," *Journal of Universal Computer Science*, vol. 21, pp. 248–267, 2015.
- [12] J. H. Lee, R. Phaal, and S.-H. Lee, "An integrated service-device-technology roadmap for smart city development," *Technological Forecasting and Social Change*, vol. 80, no. 2, pp. 286–306, 2013.
- [13] O. Söderström, T. Paasche, and F. Klauser, "Smart cities as corporate storytelling," *City*, vol. 18, no. 3, pp. 307–320, 2014.
- [14] Z. Li, Z.-Y. Tang, and X. Z., "Mapping the knowledge domain of smart-city research: a bibliometric and scientometric analysis," *Sustainability*, vol. 11, pp. 1–28, 2019.

- [15] K. Rónán, "E-regulation and the rule of law: smart government, institutional information infrastructures, and fundamental values," *Information Polity*, vol. 21, pp. 77–98, 2016.
- [16] A. Nada, A. Yasmeen, A. Manal et al., "A unified e-government service an essential component of a smart city," *International Advanced Research Journal in Science, Engineering and Technology*, vol. 7, pp. 1–10, 2017.
- [17] Z. Lv, X. Li, W. Wang et al., "Government affairs service platform for smart city," *Future Generation Computer Systems*, vol. 3, pp. 2–33, 2017.
- [18] H. Cui and P. Guo, *Smart Governance* pp. 13–14, Tsinghua University Press, Beijing, China, 1st edition, 2017.
- [19] Y. K. Wang, "Intelligent governance and intelligent service: creating a new mode of government service," *Information Construction*, vol. 01, pp. 15–18, 2017.
- [20] X. Y. Nie, "Smart governance: a new way of government governance," 2015, [http://www.cac.gov.cn/2015-11/09/c\\_1117077106.htm](http://www.cac.gov.cn/2015-11/09/c_1117077106.htm).
- [21] Y. Gao, "Smart governance: a new choice of government governance in the internet era," 2014, <https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD&dbname=CMFD2014&filename=1014203419.nh>.
- [22] S. Mills, S. Lucas, L. Irakliotis et al., *Demystifying Big Data: A Practical Guide to Transforming the Business of Government*, Tech America Foundation, Washington, DC, USA, 2012.
- [23] H. Huang, "Governance of data flow—on the theoretical evolution and framework of government data governance," *Nanjing Social Sciences*, vol. 2, pp. 53–62, 2018.
- [24] Y.-K. Xia, "Selection of data governance model of grass-roots government: construction and enlightenment of wuzhong model," *E-Government*, vol. 2, pp. 17–26, 2019.
- [25] L. Ma, "Big data governance: are local governments ready," *E-Government*, vol. 1, pp. 77–86, 2017.
- [26] H.-B. Tan and Q.-G. Meng, "Government 3.0: government governance innovation in the era of big data," *Academic Research*, vol. 12, pp. 57–61, 2018.
- [27] G. Carayannopoulos, "Whole of government: the solution to managing crises?" *Australian Journal of Public Administration*, vol. 76, no. 2, pp. 251–265, 2017.
- [28] G.-W. Zhu, "From new public administration to integral governance," *Chinese Administration*, vol. 10, pp. 52–58, 2008.
- [29] X.-G. Hu and B.-Y. Tang, "Holistic governance: a new paradigm of public management," *Journal of Central China Normal University*, vol. 01, pp. 11–15, 2010.
- [30] M. Asgarkhani, "Digital government and its effectiveness in public management reform," *Public Management Review*, vol. 7, no. 3, pp. 465–487, 2005.
- [31] M. De Claudio and C. Matthew, "Dimensions and models of contemporary public space management in England," *Journal of Environmental Planning and Management*, vol. 52, pp. 111–129, 2009.
- [32] S.-L. Wang, "Analysis framework of urban management transformation and urban governance," *Chinese Administration*, vol. 12, pp. 97–101, 2006.
- [33] S. B. Letaifa, "How to strategize smart cities: revealing the smart model," *Journal of Business Research*, vol. 68, pp. 1414–1419, 2015.
- [34] P.-B. Wu and C.-L. Lin, "Open governance innovation model of smart city: practice and enlightenment of EU and Korea," *Soft Science in China*, vol. 5, pp. 55–66, 2016.
- [35] G. Stephen and A. William, *Networked Governance: The New Shape of the Public Sector*, Peking University press, Beijing, China, 2008.
- [36] P. T. Decher, "Presidential address: false choices, policy framing, and the promise of "big data," *Journal of Policy Analysis and Management*, vol. 33, pp. 252–262, 2014.
- [37] K. W. Deutsch, *The Nerves of Government: Models of Political Communication and Control*, pp. 183–199, New York Press, New York, NY, USA, 1969.
- [38] D.-H. Shin, "Demystifying big data: anatomy of big data developmental process," *Telecommunications Policy*, vol. 40, no. 9, pp. 837–854, 2016.
- [39] J. E. Fountain, *Building the Virtual State: Information Technology and Institutional Change* pp. 69–70, Renmin University Press, Beijing, China, 1st edition, 2017.
- [40] S. Hu, *Selected Materials of Hu Shi's Philosophy*, East China Normal University Press, Shanghai, China, 1981.
- [41] M. Meijer, "Governing the smart city: a review of the literature on smart urban governance," *International Review of Administrative Sciences*, vol. 82, 2015.

## Research Article

# Research on the Impact of Green Finance and Fintech in Smart City

Zheng He,<sup>1</sup> Zhengkai Liu,<sup>2</sup> Hui Wu ,<sup>3</sup> Xiaomin Gu ,<sup>3</sup> Yuanjun Zhao,<sup>4</sup>  
and Xiaoguang Yue<sup>5</sup>

<sup>1</sup>School of Economics & Management, Shanghai Maritime University, Shanghai 201306, China

<sup>2</sup>International College, Renmin University of China, Beijing 100872, China

<sup>3</sup>School of Financial Technology, Shanghai Lixin University of Accounting and Finance, Shanghai 201209, China

<sup>4</sup>School of Business Administration, Shanghai Lixin University of Accounting and Finance, Shanghai 201209, China

<sup>5</sup>Department of Computer Science and Engineering, School of Sciences, European University Cyprus, 1516 Nicosia, Cyprus

Correspondence should be addressed to Hui Wu; wuhui@lixin.edu.cn

Received 15 October 2020; Revised 25 November 2020; Accepted 30 November 2020; Published 8 December 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Zheng He et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The green development level reflected in the green finance index and the evaluation of the degree of green development in smart cities have important practical effects on economic transformation. For promoting the transformation and upgrading of finance, we select 2013–2019 data and construct a distributed lag model to analyse the important role played by green finance and financial technology in the construction of smart cities. In the paper, we find green finance promotes the construction of smart cities, and only it will not appear until nine months late. Financial technology has an opposite impact on the construction of smart cities, this is mainly because the research and marketization of financial technology are higher, and the cost is also higher. This influence will appear in the next 14 months.

## 1. Introduction

The severe destruction of the global ecological environment is accompanied by the rapid development of the global economy. Therefore, many countries attach great importance to achieving “green growth” of their economies while vigorously developing their economies. The green upgrading of traditional industries and the growth of strategic green emerging industries require the effective empowerment of the financial sector, and the supply of green physical industries by the financial sector is a significant manifestation of green finance. Finance is the core of the modern economy. If we want to promote the economy, we must develop high-quality finance and use finance to avoid economic uncertainty. For China’s current development, green finance is the way to circumvent this uncertainty and is an important driving force for promoting supply-side structural reforms. The green development level reflected in the green finance index and the evaluation of the degree of green development in a region have important practical effects on economic transformation, reducing investment risks, and

helping traditional enterprises improve their sustainable development capabilities.

Fintech refers to financial innovation brought by new technologies. It can create new business models, applications, processes, or products, which has a significant impact on the way financial markets, financial institutions, or financial services are provided. The business scenarios where technology and finance are combined mainly include financial supervision, payment settlement, financing products and services, insurance, smart investment advisory, and energy trading. The underlying technologies include artificial intelligence, blockchain, cloud computing, big data, and the Internet of Things. In the global sustainable financial development process, financial institutions and fintech companies in various countries in the world, especially in Europe, the United States, and China, have tried to integrate fintech and financial technology by actively using blockchain, artificial intelligence, big data, and the Internet of Things, combining green finance to carry out green finance technology exploration and practice.

Smart cities are inseparable from the joint construction of all sectors of society. In the financial field, the close cooperation of the banking industry is required. With China's rapid economic and social development, the production and life of urban residents can no longer be separated from financial activities, and letting the development of financial technology benefit each inhabitant is an important part of the construction of a smart city. Correspondingly, in the era of the mobile Internet, technological means have fundamentally restructured people's behaviour habits, and banks must also change the financial services they provide in time. The construction of smart cities has been closely integrated with the development of banks. The People's Bank of China has repeatedly called on commercial banks to continuously strengthen the innovative use of financial technology to inject new vitality into urban construction. Local governments have also actively sought to cooperate with banks. The cooperation between the government and the bank has made great progress in various fields such as people's livelihood, education, and public utilities.

The specific contributions of this paper include the following:

- (1) A literature survey about various existing distributed lag models and analysis of their advantages and disadvantages.
- (2) An effective distributed lag model for green finance and fintech in the smart city is proposed.
- (3) It can be seen that there is a relatively obvious correlation between the two in the seventh lag and the twelfth lag, indicating that the two models are affected to some extent by other common variables.

The rest of this paper is organized as follows: Section 2 discusses related work, followed by the empirical analysis in Section 3. Estimation of parameters for smart cities by green finance is discussed in Section 4. Section 5 shows the discussion of the distributed lag model empirical results. Section 6 concludes the paper with a summary and future research directions.

## 2. Related Work

In foreign studies, green finance is also known as environmental finance, which is a definition of financial greening derived from environmental issues and the need to promote sustainable development. From an economic point of view, environmental issues have significant externality characteristics, which can produce positive externalities and negative externalities. Related theories include externalities, Coase's theorem, government intervention, and Pareto improvement. The development of green finance has a driving role in environmental protection, energy conservation and emission reduction, and sustainable development, so it will bring good external effects. However, good external effects directly lead to higher costs for financial institutions to implement green finance, which ultimately leads to a lack of incentives and insufficient effective supply

for the private sector to implement green finance, and it is almost impossible to achieve Pareto optimal asset allocation.

Green finance is to support sustainable development, that is, sustainable finance; it is a path of innovation in finance in environmental protection, and it is a link between the financial industry and the environmental industry [1]. Green finance is a financial tool based on market research to improve environmental quality and transfer environmental risks. As a method and means to promote sustainable development, green financial services have become the mainstream trend of financial development. Although the concept of green finance is widely recognized, it lacks a complete evaluation system, which makes it slow to develop. Some scholars have established evaluation index systems to analyse the performance problems of bank green service channels in operations and use real data from well-known banks. Based on the evaluation and analysis of their sustainable development concept and specific practices, they constructed a five-dimensional index evaluation system. Through a series of studies, it was found that although the cost of implementing the "equatorial principle" is higher, the credit risk generated by loans is lower [2-4].

The international understanding of green finance can be roughly divided into two categories: the first category regards green finance as investment and financing activities that improve the environment, that is, green investment. The starting point of this understanding is to "make up the investment gap" [2]. Another type of understanding directly incorporates environmental factors into the field of financial investment, including changes in environmental costs, changes in risk coefficients, and changes in returns and returns caused by engaging in business. It is believed that the development of green finance is based on the mutual constraints between finance and the environment [5]. For example, green finance incorporates environmental factors into financial services, such as bank investment or reception decisions, postloan monitoring, and participation in environmental assessment during risk management [6]. The above two categories are both different and related.

Over time, overseas scholars have also elaborated a series of issues in the development of green finance. Further analysis shows that the return on investment of such funds is longer with the cycle, and the overall investment income is not worse than other mutual funds.

Fintech is a modern concept and so far, no unified definition has been formed. However, in 1993, the term "fintech" appeared in the relevant report of Citibank. Attempts to enable technology to finance have gradually begun, and it has been intensified with the increasing enrichment of Internet financial practices and the continuous upgrading of the transformation of the financial sector. It can be said that fintech has quietly integrated into people's lives before society has realized it. The Financial Stability Board (FSB) believes that fintech refers to technological innovation that can significantly affect financial markets, institutions, and services and is mainly accomplished through innovative business models, application methods, process methods, and product types.

The definition of fintech has been gradually clarified in continuous research. Most scholars define fintech as a technological solution that can realize the transformation of the financial service industry. They believe that fintech is a new industry that uses technological innovation to improve the quality of financial services. Science and technology are a disruptive emerging financial industry. It mainly uses digital and information technology to achieve innovation in the existing economic model [7, 8]. In summary, fintech refers to financial innovation driven by technological innovation. Through the deep integration of finance and technology, big data, blockchain, artificial intelligence, and cloud computing are the main technical support, which has changed financial business processes. And business models challenge traditional financial thinking. Fintech has given impetus to various industries in finance-related fields by expanding the service boundaries of financial institutions and improving service methods while increasing financial efficiency and while minimizing risks [9–11]. However, fintech cannot be simply understood as the superposition of the two. It is more a process of gradual integration and deepening of emerging technologies and financial services to continuously promote the transformation and upgrading of traditional financial services.

The definition of “smart city” comes from the vision of “Smart Earth.” IBM released the theme report of “Smart Earth: Next Generation Leadership Agenda” in New York in 2008. The report mentioned “Smart Earth,” which actually refers to the full use of IT technology and sensors in various industries to embed in hospitals, power grids, railways, bridges, tunnels, highways, buildings, and other objects in the corners of the world, connect various objects to form the Internet of Things, then integrate the Internet of Things with the Internet, and finally use supercomputers and cloud computing to integrate the Internet of Things.

The construction of a smart city is not a process with clear boundaries, so there is no unified interpretation of its concept. The concept of early smart cities is mainly related to the operational functions, management, and planning efficiency of smart technology solutions in energy, transportation, physical infrastructure, distribution and communication networks, economic development, and service delivery. With the development of smart cities, more and more emphasis is placed on the use of information technology to achieve the purpose of urbanization, and a definition of a focused and operable smart city is gradually given. Invest in social capital, traditional, and modern communications infrastructure to promote sustainable economic growth and high-quality living, and wise management of natural resources. Emphasize the role of information technology. Think of smart cities as transforming productivity, to achieve the prosperity of the city and benefit the people [12–15]. Some scholars have begun to define smart cities from the perspective of the Internet. They believe that smart cities are a complex system containing multiple visions [16, 17]. Recently, with the development of information technology, more and more foreign scholars have begun to focus on sustainable development incorporating into the concept of smart cities, emphasizing the

improvement of environmental issues through technological development.

In addition, scholars have also made related research on smart city assessment. From different perspectives, they built innovative cities, regional green cities, quality of life, digital government, and so forth as evaluation indicators and carried out evaluation work on smart cities in different countries around the world [18] to promote the sustainable development of smart cities. Later, scholars carried out evaluations of smart city performance, proposed models for measuring smart city performance, and conducted empirical research.

Fintech itself is based on “science and technology,” its purpose is to promote the transformation and upgrading of finance, and it is an important force to promote the development of green finance [19]. This study explores the impact of finance on the construction of smart cities from an innovation-driven perspective and analyses the important role played by green finance and financial technology in the construction of smart cities.

### 3. Empirical Analysis

*3.1. Data Source.* The starting year of the time series data selected in this article is 2013 because 2013 is called “the first year of Internet finance.” In the study of green finance, fintech, and smart cities, this article selected seven-year data from 2013 to 2019, which can more clearly rely on the latest year’s data to understand the specific impact between the three [20]. The data are constructed through an index system. Green finance indicators include seven aspects: green credit, green bonds, green development funds, green insurance, green stocks, green investment, and carbon finance. The city index is constructed from four perspectives: smart technology, smart economy, smart society, and smart environment.

*3.2. Model Setting and Identification.* Since the independent and explanatory variables involved in this study are time series data, the structure of a commonly used linear regression model is as follows:

$$Y_t = C + \sum_{i=1}^k \beta_i X_{it} + \varepsilon_t, \quad (1)$$

in which  $Y_t$  is the time series of the dependent variable,  $X_{it}$  is the time series of the  $i$ -th independent variable,  $\varepsilon_t$  is the random disturbance term (disturbance term), and  $C$  is the constant term. In order to use the least square method (ordinary least square) for estimation, it is usually assumed that the random perturbation term and independent variable are independent of each other, and the random perturbation term is independent and identically distributed. However, the above model itself and the basic assumptions about the random perturbations have three problems: they ignore the lag effect of the independent variables, cannot handle the autocorrelation of the perturbations, and are misleadingly related [21].



In view of the above problems, this article considers the use of the dynamic regression model (dynamic regression) and introduces the lag term in the initial dynamic regression model [22]. According to the different representations of the lag term, there are two kinds of dynamic regression models: Linear Distributed Lag Model (LDL) and Rational Distributed Lag Model (RDL) [23]. Since the actual data has only a limited number of data samples, it is not possible to directly estimate the parameters of the LDL model. So this research uses the RDL model; the specific expression is as follows:

$$Y_t = C + \frac{\omega(B)B^b}{\delta(B)}X_t + \varepsilon_t = C + \frac{(\omega_0 + \omega_1 B + \dots + \omega_h B^h)B^b}{1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_r B^r}X_t + \varepsilon_t. \quad (2)$$

The RDL model can describe more complex changes in impulse response weights through the combination of multiple backshift operators, thereby better fitting the data, in which  $B^b$  is used to reflect the dead time, that is, the time interval when the independent variable starts to affect the dependent variable;  $\delta(B)$  represents the Decay Pattern; and  $\omega(B)$  reflects the absence of the impulse response weight.

This article will use the Box–Jenkins dynamic model construction strategy, namely, model identification, parameter estimation, and model diagnosis [24]. When the Box–Jenkins dynamic model construction strategy is officially started, according to the actual operation experience, the following preprocessing tasks will be performed on the original data in order, which will help the subsequent modelling process to proceed smoothly [25].

- (1) Perform differential processing on the independent and dependent variable sequences and convert them into stationary sequences  $\{X'_t\}$  and  $\{Y'_t\}$ .
- (2) Prewrite independent variable sequence to obtain the perturbation sequence  $\{a_t\}$  of the autoregressive moving average model (hereinafter referred to as ARMA model) of the independent variable sequence, which is a white noise process; its ARMA model is as follows:

$$a_t = C + \frac{\varphi(B)}{\theta(B)}X'_t. \quad (3)$$

- (3) Apply the filter expression obtained in (2) to the dependent variable sequence to obtain another sequence of perturbed terms. The process is as follows:

$$\beta_t = \frac{\varphi(B)}{\theta(B)}Y'_t. \quad (4)$$

- (4) Calculate the cross-correlation function of the sequences  $x$  and  $y$  (hereinafter abbreviated as CCF), and perform a Ljung–Box population test on the given number of lag terms  $k$ . If the null hypothesis is rejected, explain the first  $k$  lags of the independent

variable sequence. At least one of the terms has a significant nonzero correlation with the dependent variable; that is, the independent variable has explanatory power to the dependent variable and should be added to the RDL model.

Firstly, clarify three key parameters in the polynomial ratio form of the impulse response function—the molecular polynomial order  $h$ , the dead time  $b$ , and the denominator polynomial order  $r$ .

Secondly, the ARIMA model expression of the perturbation term is as follows:

$$\varepsilon_t = \frac{\Theta(B^s)\theta(B)}{\Phi(B^s)\varphi(B)}\nabla_s^D \nabla^d a_t, \quad (5)$$

in which  $a_t$  is a normally distributed white noise with a mean of 0 [26].

This paper chooses the linear transfer function method (the LTF method) to complete the entire model identification task. The LTF method requires the establishment of the following free-form distribution lag model (the FFDL model):

$$Y_t = C + v_0 X_t + v_1 X_{t-1} + \dots + v_k X_{t-k} + \frac{1}{(1 - \varphi_s B^s)(1 - \varphi_1 B)} a_t, \quad (6)$$

in which  $k$  is the number of lag effect terms chosen arbitrarily, generally the values that can be considered are  $-5$ ,  $10$ ,  $15$ , and  $20$ , and  $k$  is selected in this paper; for the disturbance term, this paper chooses a low-order autoregressive model (the AR model) to improve the accuracy of impulse response weight estimates. In the presence of seasonal effects, the AR (1) can usually be considered [27].

**3.2.1. Model Estimation.** After the model recognition task is completed, the following complete model expression can be obtained:

$$Y_t = C + \frac{\omega(B)B^b}{\delta(B)}X_t + \frac{\Theta(B^s)\theta(B)}{\Phi(B^s)\varphi(B)}\nabla_s^D \nabla^d a_t. \quad (7)$$

**3.2.2. Model Diagnosis.** Then, perform a maximum natural estimation on (6). It should be noted that both  $\{X_t\}$  and  $\{Y_t\}$  are sequences that satisfy weak stationary after performing the difference.

**3.2.3. Model Comparison.** In the process of model identification, there may be multiple candidate models. If multiple models are diagnosed through the model, they will face the problem of choosing between multiple models. There are two commonly used criteria for model comparison: Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The definition formula of AIC is as follows:

$$\text{AIC} = n \log(\hat{\sigma}^2) + 2M, \quad (8)$$

in which  $M$  is the number of parameters in the model,  $\hat{\sigma}^2$  is the maximum likelihood estimate of the variance of the disturbance term, and  $n$  is the amount of data. The definition formula of BIC is as follows:

$$\text{BIC} = n \log(\hat{\sigma}^2) + \log(n)M. \quad (9)$$

From the above definitions, when the amount of data is large ( $n > 10$ ), compared with the AIC criterion, the BIC criterion will impose a larger penalty term on the amount of model parameters. According to the Principle of Parsimony of model selection (Principle of Parsimony), this article will use BIC as the selection criterion—the model with the lowest BIC value is optimal.

**3.2.4. Construct LTF Model and Specify RDL Model Parameters.** Specify  $K$  as 20 and the perturbation term is approximated by a low-order AR model. Figures 1 and 2 show the changes in the impulse response weights of the two independent variables.

As can be seen from Figure 1, the three key parameters of the green financial independent variables in the RDL model are  $b = 4, r = 0, h = 0$ . The impulse response weight only slightly exceeds the threshold. As can be seen from Figure 2, the parameters of the RDL model part of the fintech independent variable are  $b = 14, r = 0, h = 0 + 1 - 1 = 0$ .

Figure 3 shows the residuals SACF and SPACF exhibit white noise characteristics, so the disturbance term partially retains the AR (1) model.

## 4. Results

### 4.1. Model Estimation

**4.1.1. Estimation of Parameters for Smart Cities by Green Finance.** According to the parameters specified in the model identification stage, the parameters of the RDL model are estimated based on the maximum likelihood estimation method. The results are shown in Table 1.

As can be seen from Table 1, the seasonal part of the disturbance term is not significant, and the molecular polynomial parameter of the independent variable part of RDL is not significant, so try to specify another set of alternative parameters:  $b = 9, r = 0, h = 1 + 0 - 1 = 0$ . The estimated results of the revised RDL model are shown in Table 2. It can be seen from this that each parameter passed the significance test.

According to the parameters specified in the model recognition stage, the parameters of the RDL model are estimated based on the maximum likelihood estimation method. The results are shown in Table 3. As can be seen from the figure, the seasonal part of the disturbance term is not significant, and the parameters of the fintech independent variable part are not significant, so an alternative RDL model is selected. The estimated results of the revised RDL model are shown in Table 4, which shows that each parameter passed the significance test.

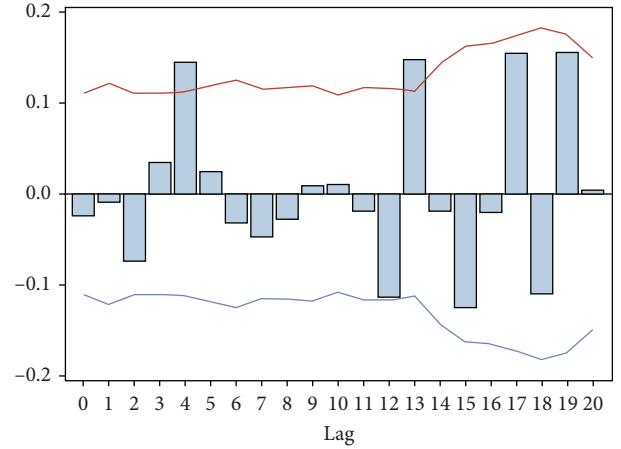


FIGURE 1: Impulse response weights for green finance.

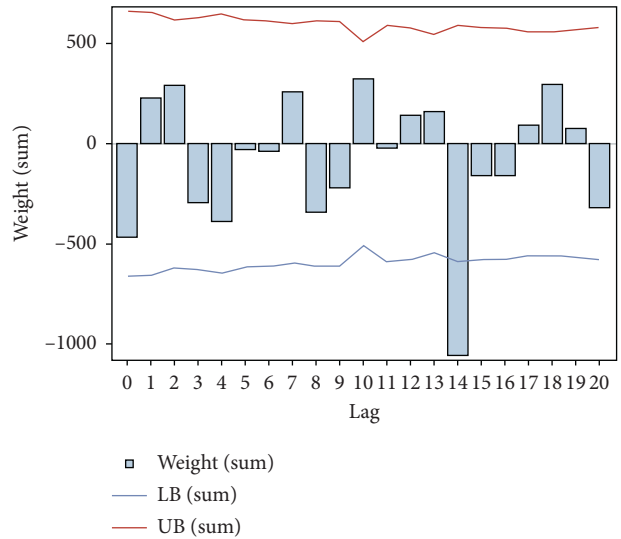


FIGURE 2: Impulse response weight map of fintech.

**4.2. Model Diagnosis and Analysis.** The first step is to perform residual diagnosis on the RDL model. As can be seen from Figures 4 and 5, the residuals of the modified RDL model can pass the autocorrelation and normality tests.

The second step is to perform residual diagnosis on the independent variable ARIMA model. The results are shown in Table 5 and Figure 6. Table 5 shows that each parameter passed the significance test, and it can be seen from Figure 6 that IACF and PACF show that the residuals pass the autocorrelation test, but they are not white noise.

In the third step, a Ljung–Box test is performed on the two residual sequences CCF, and the result is shown in Figure 7. It can be seen that there is a relatively obvious correlation between the two in the seventh lag and the twelfth lag, indicating that the two models are affected to some extent by other common variables.

From the model estimation and diagnosis results, we can know that green finance has a significant positive impact on the construction of smart cities, and this impact will occur after nine months. Fintech has a significant negative impact

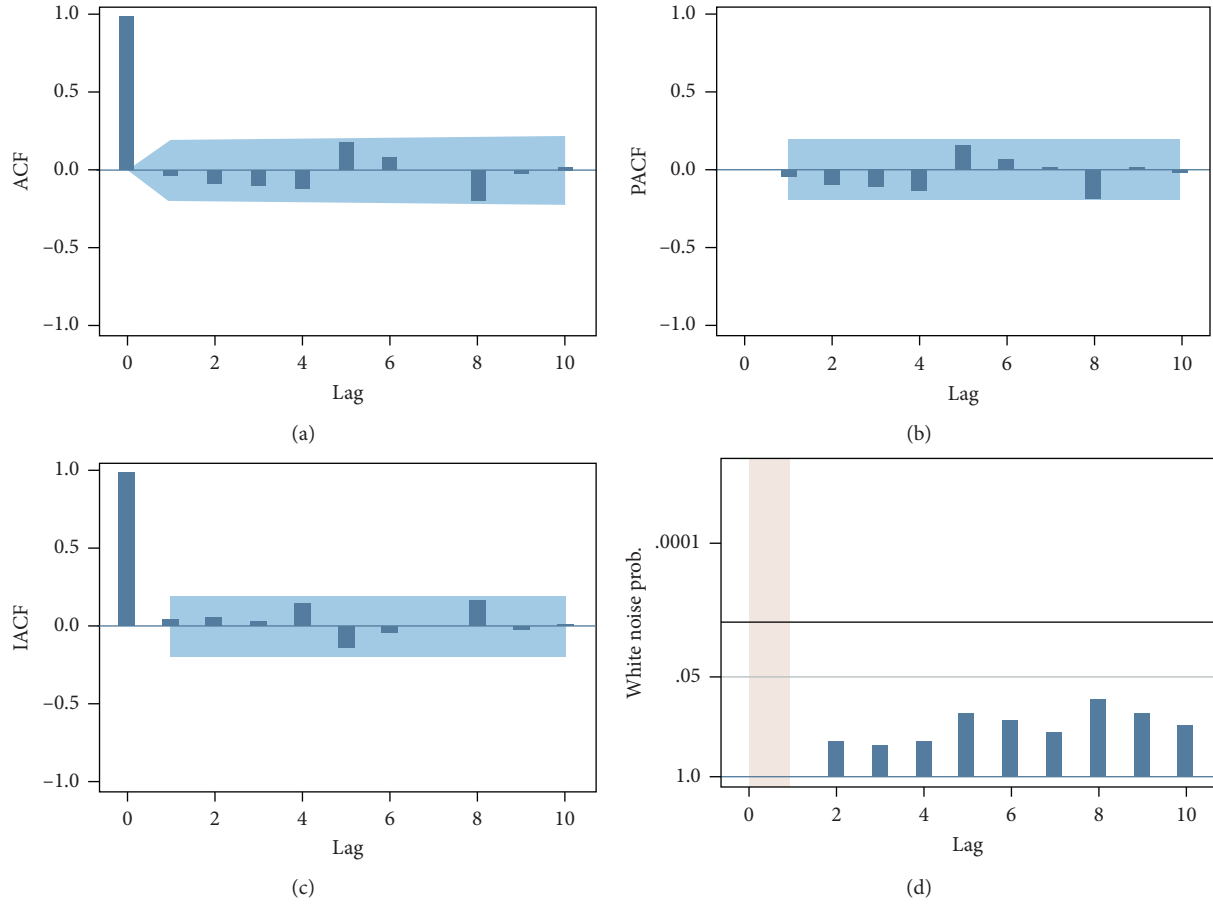


FIGURE 3: Autoregressive plots of residual error of LTF model.

TABLE 1: Estimated results of green finance RDL model.

| Parameter | Estimate | Standard error | $T$ value | Approx-<br>$t < P < t$ | Lag |
|-----------|----------|----------------|-----------|------------------------|-----|
| MU        | 0.0031   | 0.0021         | 1.47      | 0.1409                 | 0   |
| MA1,1     | 0.2817   | 0.0919         | 3.06      | 0.0022                 | 1   |
| MA1,2     | 0.3593   | 0.0946         | 3.80      | 0.0001                 | 2   |
| MA2,1     | -0.1337  | 0.1018         | -1.31     | 0.1891                 | 11  |
| NUM1      | 0.0054   | 0.0010         | 5.08      | <0.001                 | 0   |
| NUM1,1    | -0.0013  | 0.0010         | -1.33     | 0.1823                 | 1   |

TABLE 2: Estimated results of the revised RDL model for green finance.

| Parameter | Estimate | Standard error | $T$ value | Approx-<br>$t < P < t$ | Lag |
|-----------|----------|----------------|-----------|------------------------|-----|
| MU        | 0.0029   | 0.0018         | 1.67      | 0.0940                 | 0   |
| MA1,1     | 0.3028   | 0.0905         | 3.34      | 0.0008                 | 0   |
| MA1,2     | 0.3582   | 0.0932         | 3.85      | 0.0001                 | 0   |
| NUM1      | -0.0053  | 0.0011         | 4.81      | <0.001                 | 9   |

on smart cities, and this impact occurs after 14 months. The model diagnosis results show that there are common variables in the residuals of the RDL model and ARIMA model

TABLE 3: Estimation results of fintech RDL model.

| Parameter | Estimate | Standard error | $T$ value | Approx-<br>$t < P < t$ | Lag |
|-----------|----------|----------------|-----------|------------------------|-----|
| MU        | 0.0545   | 0.0011         | 1.17      | 0.2503                 | 0   |
| AR1,1     | 0.0598   | 0.0987         | -5.06     | 0.0564                 | 0   |
| AR2,1     | 0.0569   | 0.0945         | -1.65     | 0.0641                 | 0   |
| NUM1      | 0.0065   | 0.0965         | 4.12      | <0.001                 | 14  |
| NUM2      | 0.0058   | 0.0354         | -1.25     | <0.001                 | 6   |
| NUM1,1    | 0.0965   | 0.0569         | 0.59      | <0.001                 | 6   |

TABLE 4: RDL model estimation results.

| Parameter | Estimate | Standard error | $T$ value | Approx-<br>$t < P < t$ | Lag |
|-----------|----------|----------------|-----------|------------------------|-----|
| MU        | 0.0596   | 0.0023         | 1.19      | 0.2336                 | 0   |
| AR1,1     | -0.0632  | 0.0895         | -6.03     | <0.001                 | 0   |
| NUM1      | 0.0782   | 0.0652         | 2.48      | 0.0130                 | 14  |

constructed in this paper. Due to the influence of common variables, the results of their specific parameter estimation may not be stable. Figure 7 is CCF significance test chart of two residual sequences.

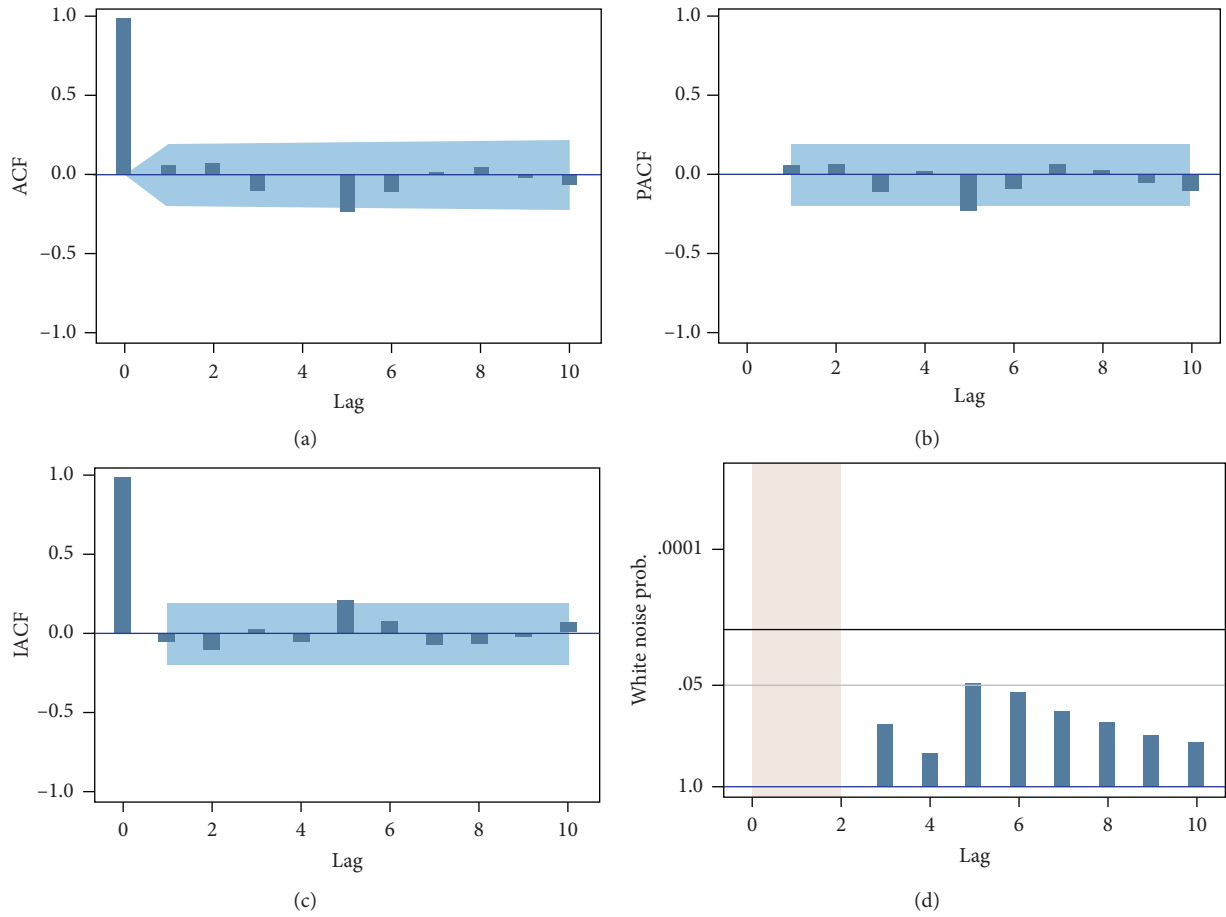


FIGURE 4: Residual normality test chart of RDL model after green finance correction.

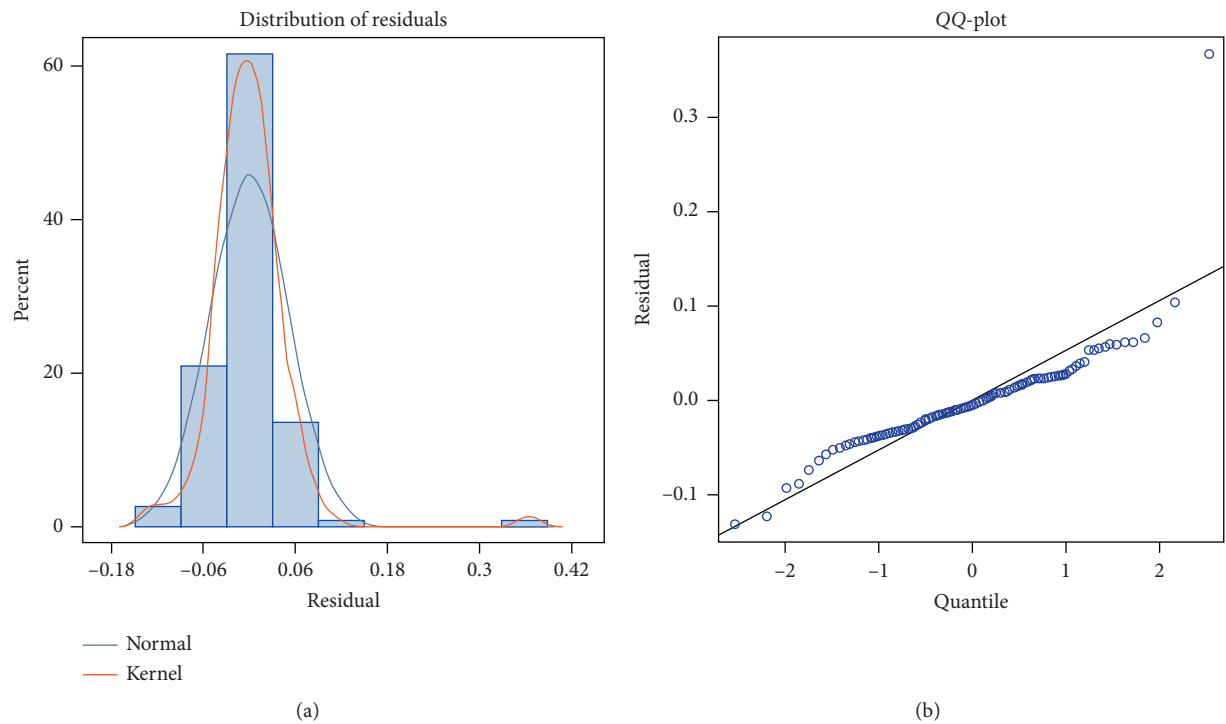


FIGURE 5: Autocorrelation diagnosis diagrams of the residual error of the RDL model after green finance correction.

TABLE 5: Maximum likelihood estimation results of the fintech ARIMA model.

| Parameter | Estimate | Standard error | $T$ value | Approx- $t < P < t$ | Lag |
|-----------|----------|----------------|-----------|---------------------|-----|
| MU        | -0.0642  | 0.0865         | -0.08     | 0.9358              | 0   |
| AR1,1     | -0.5653  | 0.0824         | -2.60     | 0.0082              | 1   |
| AR1,2     | -0.9252  | 0.0963         | -4.26     | <0.001              | 2   |
| AR1,3     | -0.4545  | 0.0918         | -6.35     | <0.001              | 3   |
| AR1,4     | -0.2545  | 0.0854         | -2.26     | 0.0235              | 4   |
| AR2,1     | -0.2224  | 0.0871         | -7.21     | <0.001              | 11  |
| AR2,2     | -0.7554  | 0.1025         | -4.26     | <0.001              | 22  |
| AR2,3     | -0.7893  | 0.0965         | -2.60     | 0.0093              | 33  |

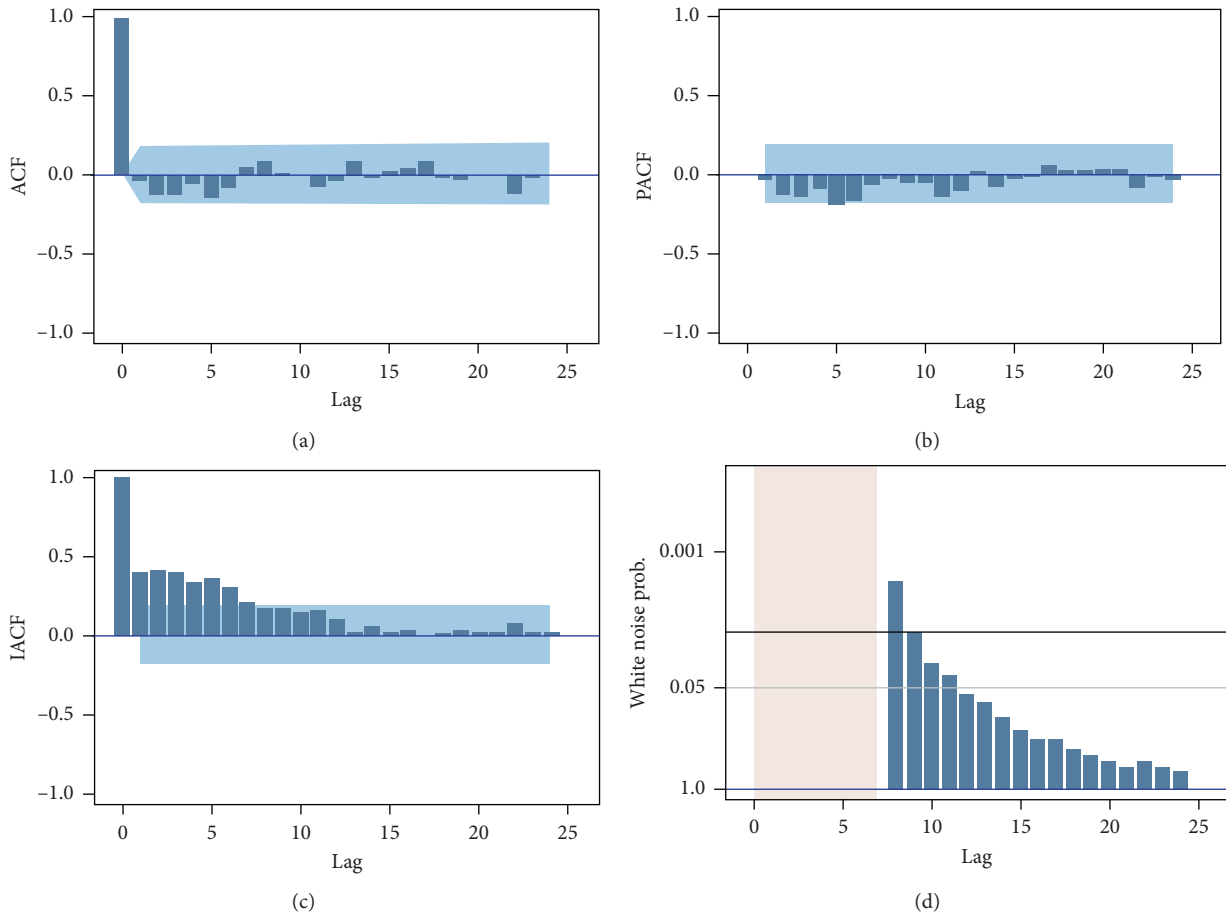


FIGURE 6: Autocorrelation test chart of fintech ARIMA model residuals.

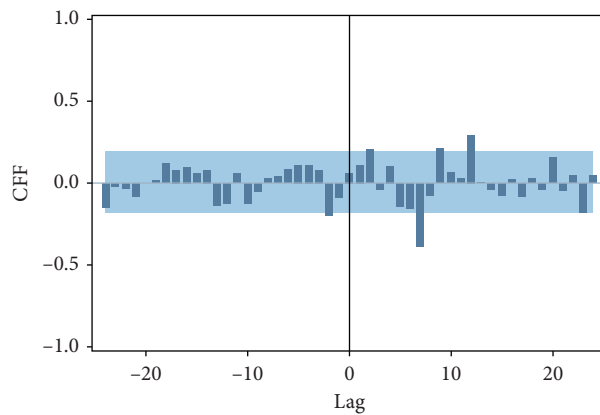


FIGURE 7: CCF significance test chart of two residual sequences.

## 5. Discussion

The first is to innovate financial institutions and vigorously develop technology banks, technology microcredit companies, and financial leasing companies. On the premise of following the law of development of technology innovation, expand diversified financing channels, and, at the same time, use mechanism innovation as a driving force to strengthen internal incentives. Improve fair and objective assessment mechanisms and improve protection systems such as risk compensation. The second is to innovate financial products, efficiently use differential risk tolerance to support scientific and technological credit, promote high-yield bond and equity bond financing methods, and comprehensively promote intellectual property pledge financing and repayment. Innovation in financing methods: the third is innovative financing methods, breaking through the traditional service model of indirect financing, financial team and equity, and venture capital investment and loan linkage; the fourth is focusing on the development opportunities of Internet financial technology innovation, fostering "Internet thinking," and actively promoting "networks" "Process" model; creating Internet banking, high-quality service platforms, electronic banking, online revolving loans, and other IT banks; effectively solving the "gap" in funds faced by high-tech projects; promoting the output and diffusion of technological innovation; and promoting the development of information technology in smart cities.

The core of the development of green finance lies in financial intermediaries. The market's actual demand for green finance requires financial institutions to continuously improve their awareness of social responsibility and environmental protection, and carry out targeted green finance promotion and education for internal staff. When making business decisions, we must coordinate the relationship between social responsibility and self-interest. We must always consider environmental protection and sociality in daily operations. Strict environmental protection social responsibility runs through all daily business activities. The financial community must maintain the concept of developing green finance at all times. It must take a dialectical and unified view of the social responsibility and self-realized benefits of enterprises when they conduct green financial business and promote the development of green finance as a long-term strategy to operate, financial institutions. The government should actively carry out green finance publicity work, publicize the government's ideas and measures in developing green finance, and explain the positive changes that green finance development will bring to the lives and work of the general public. On this basis, companies can actively participate in promoting the development of green finance.

One of the major obstacles facing the development of China's green finance is the incomplete information sharing mechanism between different institutions in green finance and the problem of information asymmetry in different sectors, especially the financial and environmental protection sectors that are closely related to green finance. The information asymmetry between the two departments lays

hidden dangers for the occurrence of financial risks. Through the application of fintech, a green financial information database can be established between departments, thereby forming a stable communication and coordination mechanism and achieving information sharing between the two parties. At the same time, regulators can also use real-time tracking and monitoring of enterprises participating in green credit through the application of fintech and update related information in real time. In addition, a mature information disclosure system should also be established through the application of fintech, the disclosure procedures for financial institutions to provide green financial services to enterprises, the role of the public and groups in supervision, and the strengthening to cooperate with each other to create a system of vertical supervision and horizontal supervision.

## 6. Conclusions

Starting from the relationship between green finance, fintech, and smart city construction, this study selects data from 2013 to 2019 and uses a distributed lag model to demonstrate it. The following conclusions are obtained: first, green finance promotes smart city construction and green credit. The promotion effect of green financial indicators composed of seven aspects including green bonds, green bonds, green development funds, green insurance, green stocks, green investment, and carbon finance on the construction of smart cities will not appear until nine months later. The honey lake fintech index has verified that it has a reverse effect on the construction of smart cities through models. This is mainly due to the high cost of fintech research, marketization, and profitability, which hinders the construction of smart cities to a certain extent. The effect of this effect will appear after 14 months.

## Data Availability

Some or all data, models, or codes generated or used during the study are available from the corresponding author by request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The funding was sponsored by the National Social Science Fund of China (Grant no. 18CGL015).

## References

- [1] M. A. White, "Environmental finance: value and risk in an age of ecology," *Business Strategy & the Environment*, vol. 5, no. 3, pp. 198–206, 1996.
- [2] A. B. Coulson, "Sustainable banking and finance: people-the financial sector and the future of the planet," *Futures*, vol. 35, no. 8, pp. 894–895, 2003.
- [3] S. Dimitriadis, A. Kouremenos, and N. Kyrezi, "Trust-based segmentation: preliminary evidence from technology-enabled

- bank channels," *International Journal of Bank Marketing*, vol. 29, no. 29, pp. 5–31, 2011.
- [4] S. C. Morrish and C. Lee, "Country of origin as a source of sustainable competitive advantage: the case for international higher education institutions in New Zealand," *Journal of Strategic Marketing*, vol. 19, no. 6, pp. 517–529, 2011.
- [5] K. Kaufer, "Social responsibility as a core business model in banking: a case study in the financial sector," *Journal of Sustainable Finance & Investment*, vol. 4, no. 1, pp. 76–89, 2014.
- [6] N. Apergis, I. Filippidis, and C. Economidou, "Financial deepening and economic growth linkages: a panel data analysis," *Review of World Economics*, vol. 143, no. 1, pp. 179–198, 2007.
- [7] C. Wright, "Global banks, the environment, and human rights: the impact of the equator principles on lending policies and practices," *Global Environmental Politics*, vol. 12, no. 1, pp. 56–77, 2012.
- [8] B. M. S. van Praag, P. Frijters, and A. Ferrer-i-Carbonell, "The anatomy of subjective well-being," *Journal of Economic Behavior & Organization*, vol. 51, no. 1, pp. 29–49, 2003.
- [9] H. M. Mueller and C. Yannelis, "The rise in student loan defaults," *Journal of Financial Economics*, vol. 131, no. 1, pp. 1–19, 2019.
- [10] N. D. Philippas and C. Avdoulas, "Financial literacy and financial well-being among generation-Z university students: evidence from Greece," *The European Journal of Finance*, vol. 26, no. 4-5, pp. 360–381, 2020.
- [11] M. Shaban, C. Girardone, and A. Sarkisyan, "cross-country variation in financial inclusion: a global perspective," *European Journal of Finance*, vol. 26, no. 4-5, pp. 319–340, 2020.
- [12] M. Swan, "Anticipating the economic benefits of blockchain," *Technology Innovation Management Review*, vol. 7, no. 10, pp. 6–13, 2017.
- [13] M. Swan and P. de Filippi, "Toward a philosophy of blockchain: a symposium: introduction," *Metaphilosophy*, vol. 48, no. 5, pp. 603–619, 2017.
- [14] Y. Yun and M. Lee, "Smart city 4.0 from the perspective of open innovation," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 5, no. 4, pp. 92–107, 2019.
- [15] R. U. Arora, "Financial sector development and smart cities: the Indian case," *Sustainable Cities and Society*, vol. 42, pp. 52–58, 2018.
- [16] Y. J. Shin and Y. Choi, "Feasibility of the fintech industry as an innovation platform for sustainable economic growth in Korea," *Sustainability*, vol. 11, no. 19, pp. 5351–5372, 2019.
- [17] P. Schulte and G. Liu, "FinTech is merging with IoT and AI to challenge banks: how entrenched interests can prepare," *The Journal of Alternative Investments*, vol. 20, no. 3, pp. 41–57, 2017.
- [18] S. Heinemann and M. Kreyling, "Paving the way towards smart cities using AI-digital automotive services research project," *ATZelectronics Worldwide*, vol. 14, no. 1-2, pp. 54–59, 2019.
- [19] W. Sun, Y. Zhao, and L. Sun, "Big data analytics for venture capital application: towards innovation performance improvement," *International Journal of Information Management*, vol. 50, no. 2, pp. 557–565, 2019.
- [20] H. Liang, J. Zou, K. Zuo, and K. Muhammad Junaid, "An improved genetic algorithm optimization fuzzy controller applied to the wellhead back pressure control system," *Mechanical Systems and Signal Processing*, vol. 142, no. 8, pp. 1–14, 2020.
- [21] L. Sun, Y. Zhao, W. Sun, and Z. Liu, "Study on supply chain strategy based on cost income model and multi-access edge computing under the background of the Internet of Things," *Neural Computing and Applications*, vol. 32, no. 19, pp. 15357–15368, 2020.
- [22] Z. Huang, X. Xu, J. Ni, H. Zhu, and C. Wang, "Multimodal representation learning for recommendation in Internet of Things," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10675–10685, 2019.
- [23] H. Liang, J. Zou, Z. Li, M. J. Khan, and Y. Lu, "Dynamic evaluation of drilling leakage risk based on fuzzy theory and PSO-SVR algorithm," *Future Generation Computer Systems*, vol. 95, no. 6, pp. 454–466, 2019.
- [24] M. Xie, H. Li, and Y. Zhao, "Blockchain financial investment based on deep learning network algorithm," *Journal of Computational and Applied Mathematics*, vol. 372, no. 7, pp. 1–11, 2020.
- [25] Z. Liu, B. Hu, B. Huang, L. Lang et al., "Decision optimization of low-carbon dual-channel supply chain of auto parts based on smart city architecture," *Complexity*, vol. 2020, no. 5, pp. 1–14, 2020.
- [26] Y. Zhang, R. Zhu, Z. Chen, J. Gao et al., "Evaluating and selecting features via information theoretic lower bounds of feature inner correlations for high-dimensional data," *European Journal of Operational Research*, vol. 6, 2020.
- [27] H. Liang, A. Xian, M. Min Mao, P. Ni et al., "A research on remote fracturing monitoring and decision-making method supporting smart city," *Sustainable Cities and Society*, vol. 62, no. 11, pp. 1–11, 2020.

## Research Article

# Application Research of Intelligent Classification Technology in Enterprise Data Classification and Gradation System

Lina Yu <sup>1</sup>, Chunwei Wang <sup>1,2</sup>, Huixian Chang <sup>1</sup>, Sheng Shen <sup>1</sup>, Fang Hou <sup>1</sup>,  
and Yingwei Li <sup>1</sup>

<sup>1</sup>School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China

<sup>2</sup>Beijing Branch, Daqing Oilfield Information Technology Company, Beijing 100043, China

Correspondence should be addressed to Yingwei Li; [lyw@ysu.edu.cn](mailto:lyw@ysu.edu.cn)

Received 22 October 2020; Revised 18 November 2020; Accepted 25 November 2020; Published 7 December 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Lina Yu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Classification and gradation system adopts different security protection schemes for different types of data by implementing classification and gradation management of data, which is an important pretechnical means for data security protection and prevention of data leakage. This paper introduces artificial intelligence classification, machine learning, and other means to learn and train enterprise documents according to the characteristics of enterprise sensitive data. The generated training model can intelligently identify and classify file streams, improving work efficiency and accuracy of classification and gradation. At the same time, the differences, advantages, and disadvantages of K-NN (K-Nearest Neighbors), DT (Decision Tree), and LinearSVC algorithms are compared. The experimental data shows that LinearSVC algorithm is applicable to high-dimensional data, with discrete, sparse data features and large number of features, which is more suitable for classification of sensitive data of enterprises.

## 1. Introduction

With the advent of Internet plus era, the status of data as a basic strategic resource has become increasingly prominent [1–3]. As network security threats become increasingly prominent, data face many security risks during storage, processing, and transmission. The challenges of management and monitoring which accompany it are increasingly severe [4, 5], so the establishment of a data classification and gradation system environment is the basic premise for the rapid development of enterprises [6, 7]. According to the results of data classification and gradation and the company's policy requirements, efficiently matching the collated data with security strategies is an important means for enterprises to improve their core competitiveness. For large enterprises, data security faces many problems and challenges. According to the international standards of information security [8], the importance of different data is different, and high-value data requires stricter protection mechanisms. Therefore, data is used as the security protection target, and the intricate enterprise data assets are divided into various categories and multiple levels according

to the classification and gradation method. According to the type and value of data, different protection strategies are formulated [9, 10], and the continuous strengthening and improvement of sensitive data security management have become more prominent and important.

At present, the state has promulgated laws and regulations such as the "Network Security Law of the People's Republic of China" [11], clearly stating that "network operators should comply with the requirements of the network security level protection system, perform the following security protection obligations to protect the network from interference, destruction, or unauthorized access, prevent network data from being leaked, stolen, or tampered," and "take measures such as data classification, important data backup, and encryption". The data combining work has the following problems. First, for the sensitive data of enterprises, due to the lack of strict management of the computer network by the managers and their lack of awareness of network security, it is easy for the computer to leak information, which causes a series of subsequent security risks to damage the computer network [12]. Second, most of the understanding of policies, regulations, and standards also



relies on manual combining. There is a possibility that the understanding and interpretation of policies, regulations, and standards may be artificially expanded or reduced. Third, at the same time, with regard to the characteristics of large data business types and large data volumes of large enterprises [13, 14], business personnel cannot quickly identify which sensitive data is based on standards, and the level of confidentiality corresponding to sensitive data is not easy to define. Therefore, establishing a classification and gradation protection catalogue for the corresponding trade secrets of large enterprises, establishing a more detailed classification and gradation process specification and implementation guide according to the business data, improving the efficiency and accuracy of classification and gradation management, providing reference for confidentiality of documents for business personnel, and forming compliance management measures are imperative [15, 16].

By integrating artificial intelligence classification technology [17], this paper carries out automatic classification and gradation of documents. This paper compares three classification algorithms and shows the feasibility and effect of LinearSVC algorithm in the classification and gradation system, so as to improve the accuracy of data classification and gradation [18].

## 2. Classification and Gradation System Architecture and Deployment

The classification and gradation management of enterprise information can enable various sensitive data information of the enterprise to be grasped in a timely, efficient, and accurate manner [19]. It is an important pretechnical means for large-scale enterprises to protect data security and prevent data leakage. According to the data privacy standards and management systems of enterprises, classification and gradation management of sensitive data of enterprises is more convenient for the formulation of data security policies and the protection of sensitive data.

Figure 1 shows the deployment of classification and gradation. As the core part of classification and gradation, AI server can predict the classification and gradation of enterprise documents through artificial intelligence classification and machine learning and provide users with a higher level of intelligent application services; the role of the load-balancing server is to distribute the user to the web server group; the file server and the database, respectively, provide storage for the file and relational data for the classification and gradation system; and the report server provides a comprehensive display for the classification and gradation information.

The DLP system continuously captures and analyzes the traffic on the network by placing a monitor at the outlet of the enterprise's external network and detects sensitive data and important traffic elements [20] through protocols such as SMTP, FTP, and HTTP to prevent the transfer of sensitive data to the outside. It can be seen from Figure 2 that the classification and gradation system integrates with the DLP (data leakage prevention) system to obtain network or terminal data leakage event information and displays related

information for organizations with data permissions on the basis of strict authority control design. The user data protection requirements can be collected from the information or security management department to provide a basis for the data security policy. At the same time, the DLP system interacts with the classification and gradation system to compare the fingerprint of the outgoing file with the fingerprint database of the classification and gradation system to predict which category the file belongs to.

## 3. Materials and Methods

Data intelligent classification mainly uses intelligent classification technology to form different categories of data to be classified [21]. As shown in Figure 3, the AI intelligent classification function is mainly divided into two modules: AI training and AI classification. The AI training module is processed by a stand-alone server. Through the full amount of learning from classification and gradation information source uploaded by users, generating the system model, and uploading it to AI classification module, AI classification will classify documents according to the model. In the AI classification module, when the user enters the classification and gradation information, the platform intelligently classifies and gives the user a classification prompt to provide a reference for the user.

*3.1. Data Preprocessing.* To be able to calculate the accuracy of the classification algorithm, before the model training, the system needs to automatically classify and annotate the data sample files and then classify the inaccurate files automatically into the correct classification through manual proofreading. The manual proofreading step is very important and has a great impact on the final accuracy. For the calibration document after proofreading, if obtaining higher quality corpus data is needed during model training, the data should be preprocessed in advance [22]. The data preprocessing flow chart is shown in Figure 4.

The original text is first cleaned. This is the last procedure to find and correct identifiable errors in the data file. It reexamines and verifies the data, removing duplicate information, correcting existing errors, and providing data consistency. After that, the word segmentation process is performed, the Chinese characters are divided into individual words, and the continuous word sequences are recombined into word sequences according to certain specifications. Finally, the stop words are removed. The words that do not contribute to the text features are roughly divided into two categories. One kind of stop words is characterized by a wide range of applications and can be found everywhere in various documents; for example, the word "company" appears in almost every document, and the characteristics of the document cannot be reflected for such words. Another kind of stop words includes a modal particle, an adverb, a preposition, and a conjunction, which usually has no clear meaning. These words will not have specific meaning until they are put into a complete sentence, such as the common "the," "in," and the like. After the data is

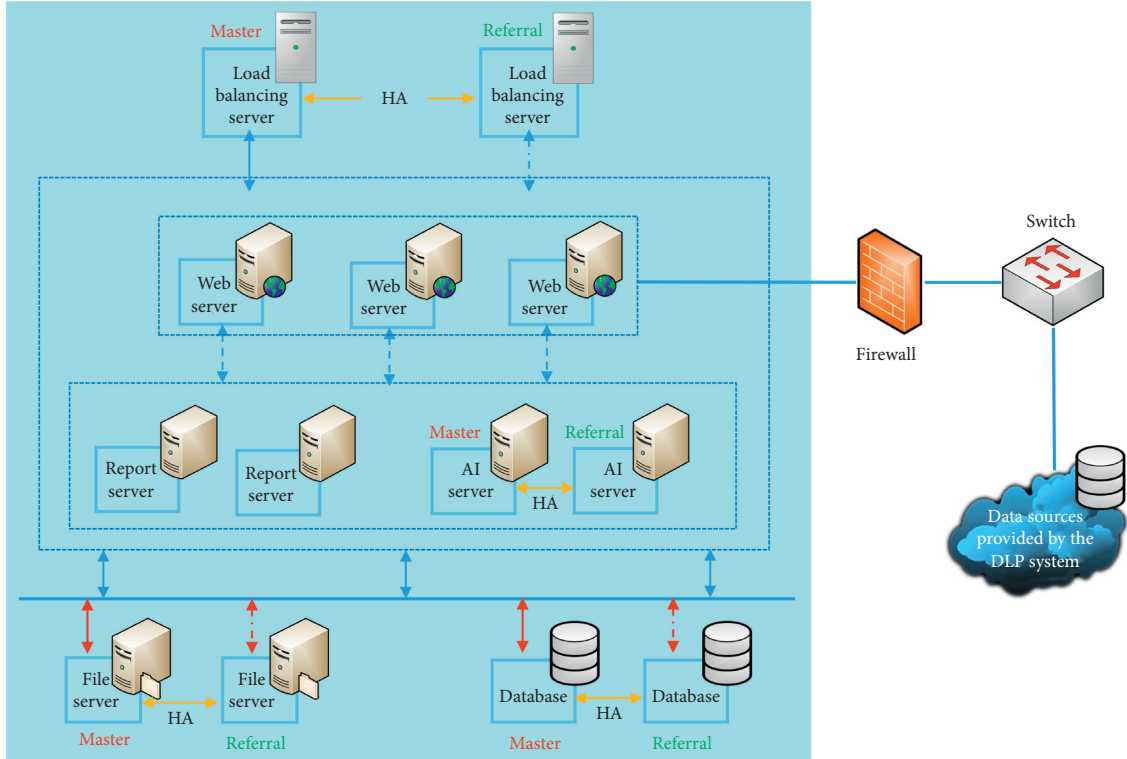


FIGURE 1: Classification and gradation system architecture diagram.

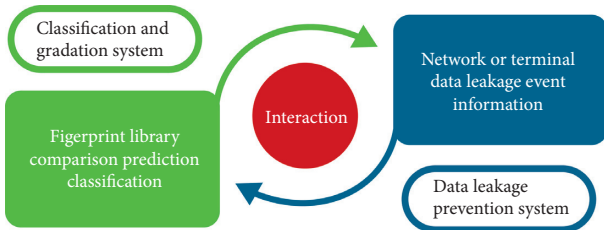


FIGURE 2: Interaction diagram of classification and gradation system and DLP system.

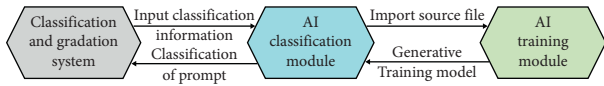


FIGURE 3: AI intelligent classification module diagram.

preprocessed, the original text is expected to have higher quality prediction data for the next model training.

**3.2. Model Training.** Dividing all training data into two parts, one for the training model is called the train set, and the other for the accuracy test of the model is called the test set. Then the TF-IDF (term frequency-inverse document frequency) [23] calculation on the two data sets is carried out. The calculation process is as follows: for the word  $t_i$  in each parsed.txt file  $d_j$ , the term frequency (TF) can be expressed as in the following equation:

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}, \quad (1)$$

where  $n_{i,j}$  is the number of occurrences of the term  $t_i$  in the document  $d_j$  and the denominator is the sum of the occurrences of all the terms in the document  $d_j$ . The main idea of IDF (inverse document frequency) is that if there are fewer documents containing the word  $t_i$ , the IDF value is larger, indicating that the word  $t_i$  has a good class discrimination ability at the level of the entire document set. IDF is expressed as follows:

$$idf_i = \log \frac{|D|}{|\{j: t_i \in d_j\}|}, \quad (2)$$

where  $|D|$  is the total number of all documents in the corpus and the denominator is the number of all documents containing the term  $t_i$ . The TF-IDF weight is actually the product of two parameters. That is,  $tf_{i,j} \times idf_i$ .

TF-IDF is a commonly used weighting technique for information retrieval and data mining. It is a statistical method used to evaluate the importance of a word to a document set or one of the documents in a corpus [24]. The importance of a word increases in proportion to the number of times it appears in the document, but at the same time it decreases in inverse proportion to the frequency of its appearance in the corpus. If in a specific type of test set document the TF-IDF value of the word  $t_i$  is high but that in other categories is very low or even 0, indicating that the word is of greater importance to this type of document and has a strong ability to classify this type of document, it can be

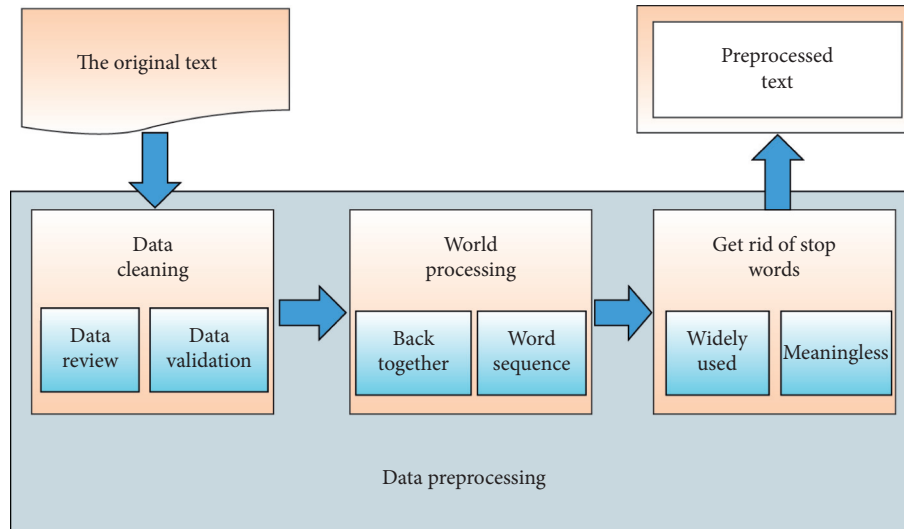


FIGURE 4: Data preprocessing flow chart.

regarded as the characteristic word of this kind of file. The specific model training process is as follows: TF-IDF calculation is performed on the words in the file data in each category of the training set; then the words with the strongest classification ability are selected as the feature words of the type of file. So, the classifier training is performed using these features. In the end, a classification model is generated. After each training of the model, the test data set needs to be used to verify the accuracy of the model. When the accuracy reaches the requirement, the model can be saved. If the higher accuracy is not achieved, it is necessary to classify the annotation data to confirm whether there is a data classification error and then adjust the parameters of the model, evaluate the model, and iterate the process until a higher accuracy is obtained. The model training process is shown in Figure 5.

**3.3. Learning Function of Classification Model.** In view of the numerous and intricate features of the documents in the enterprise, with the change of the enterprise system and the continuous increase of documents, it is sometimes necessary to make changes to the classification standards and adjust the data file categories. The AI intelligent classification model learning framework is shown in Figure 6. The framework consists of model learning services, file parsing services, and model training. With the corpus and new standard of classification reuploaded, the file parsing service is called to convert the data file into a txt file, and the corpus is initially automatically classified based on the new classification criteria. After manual proofreading, the model training module is invoked for training. The result is repeated iterations to achieve the learning function of the model.

## 4. Results and Discussion

**4.1. Several Common Classification Algorithms.** Logistic Regression (LR) [25], a generalized linear regression analysis model, is based on the Sigmoid function to deal with large-

scale data by giving the probability that the sample belongs to each category. Logistic regression algorithms are often used to solve the two classification problems; Naive Bayes (NB) [26], a method for calculating posterior probabilities from prior probabilities, requires a hypothetical premise. In the actual data classification analysis process, this premise assumption is often too idealistic, and it is not established in the actual situation. So the above two algorithms are not suitable for use in the enterprise data multiclassification system.

Decision Tree (DT) [27] creates a tree node by calculating the information gain of each attribute and selecting the attribute with the highest information gain as the test attribute of the given data set, and marks with this attribute, and then creates a separate branch for each value of the attribute and divides the sample accordingly; K-Nearest Neighbors (K-NN) [28] classification, an analogy-based learning method, works by storing all training samples in an N-dimensional model space, calculating sample files by calculating the K training samples closest to a given unknown sample using the Euclidean distance formula; LinearSVC [29] is one of the SVM (Support Vector Machine) classification algorithms. By using kernel function techniques, linearly inseparable features are mapped into high-dimensional space, so that features can be divided in high-dimensional space. Based on the limited sample information, the complexity of the model (the learning accuracy of a particular training sample) and the learning ability (the ability to identify any sample without error) maximize the maximum separation between the separate categories to achieve good classification predictions for the sample file.

**4.2. Experimental Scheme and Results.** Intelligent classification and gradation module is essentially a module of text classification. Text classification refers to the automatic classification process of the input text according to a certain categorization system by the computer through algorithms. The algorithm of this classification and gradation module is implemented by a more mature machine learning algorithm.

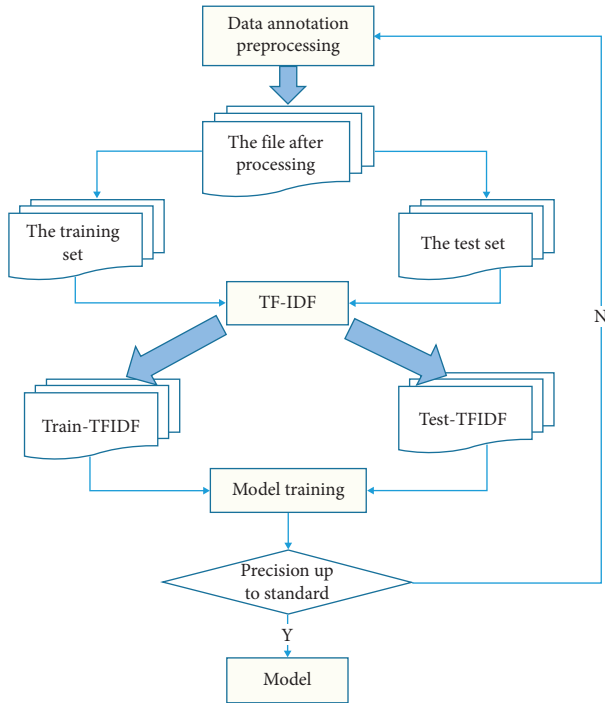


FIGURE 5: Model training flow chart.

In machine learning, there are many algorithms that can be used for text classification. After comparing the advantages and disadvantages of different algorithms, combined with the sparse and discrete features of the enterprise text data, three algorithms (Decision Tree, K-NN, and LinearSVC) that are suitable for enterprise text classification are selected for experiment.

For the experimental work of exploring the data volume of the model training on the accuracy and modeling time of the three classification algorithms, according to most enterprise systems and organizational components, the enterprise data can generally be roughly divided into 12 categories including personnel, auditing, legal affairs, material procurement, production and management, technology management, discipline inspection and supervision, maintenance of letters and visits, comprehensive office, planning, finance, international cooperation, and policy research. For the discrete, sparse, and feature-rich characteristics of the enterprise data, if there are too many training samples, some abnormal feature values will appear inside for the discrete and sparse enterprise data, which will affect the accuracy of the model. At the same time, the amount of training data should not be too small; otherwise, it will also affect the accuracy. In view of the fact that the number of training samples would affect the accuracy of the classification model, this experiment fixed the size of the sample categories into 12 to explore the changes in the accuracy and modeling time of the three classification algorithms when the amount of training data for each classification file is 40, 60, 80, 100, 120, 140, and 160. It is worth noting that, in order to avoid the influence of other factors on the experiments, all experiments were conducted on the same hardware configuration server. The results are shown in Figure 7.

Due to the characteristics that the enterprise data are discrete, sparse, and numerous, the more categories are classified, the higher the degree of feature coincidence among each category is, which makes the accuracy of the model more easily affected. In general, the enterprise documents are roughly divided into 12 categories: company personnel, finance, and so on. According to the different nature and the different system of the enterprise, the data category species will slightly change.

When exploring the influence of the number of training samples on the model accuracy and modeling time, it can be seen from Figure 7 that the modeling time increases with the increase of the number of training samples, while the accuracy of the model built by the three algorithms peaks within the range of local sample size (75–100). Therefore, as for the experimental work to explore the influence of the number of enterprise data classification types on the accuracy and modeling time of the three classification algorithms, in order to facilitate the calculation of experimental parameters, the fixed training sample size for this experiment is 100 files. Therefore, this experiment explores the changes in the accuracy and modeling time of the three classification algorithms when the number of classification types is 8, 10, 12, 14, 16, 18, and 20 in sequence for servers with the same hardware configuration. The variations of the three algorithms are displayed in Figure 8.

**4.3. Analysis of Experimental Results.** It can be seen from the data and the line graph obtained from the two experiments that, for different numbers of training samples, the model training time of each algorithm is almost the same for the three algorithms. But the training time of the Decision Tree algorithm model is generally over 1s, while the other two algorithms K-NN and LinearSVC have slightly faster speed, with the over 1s training time when the training sample size is more than 100. However, in the view of the accuracy, LinearSVC classification algorithm is superior to the other two kinds of algorithm; its accuracy can reach 95% or so. The LinearSVC classification algorithm has the highest accuracy when the number of training samples is about 100 files. For the experiments for different types of enterprise data classification, the training times of all algorithms are still similar, but the accuracy of the LinearSVC classification algorithm is still the highest, which can reach about 95%. According to the results of the two experiments, it can be seen that, when considering the time used for modeling, the three classification algorithms are not much different, but the LinearSVC classification algorithm still has a far better accuracy than the other two algorithms in the two experiments and therefore is most suitable for application in the enterprise data classification and gradation system.

## 5. Application Verification and Summary

**5.1. Analysis of Experimental Results.** Taking a petroleum enterprise as an example, the AI intelligent classification composed of the LinearSVC algorithm is applied in the classification and gradation system of the enterprise. Then,

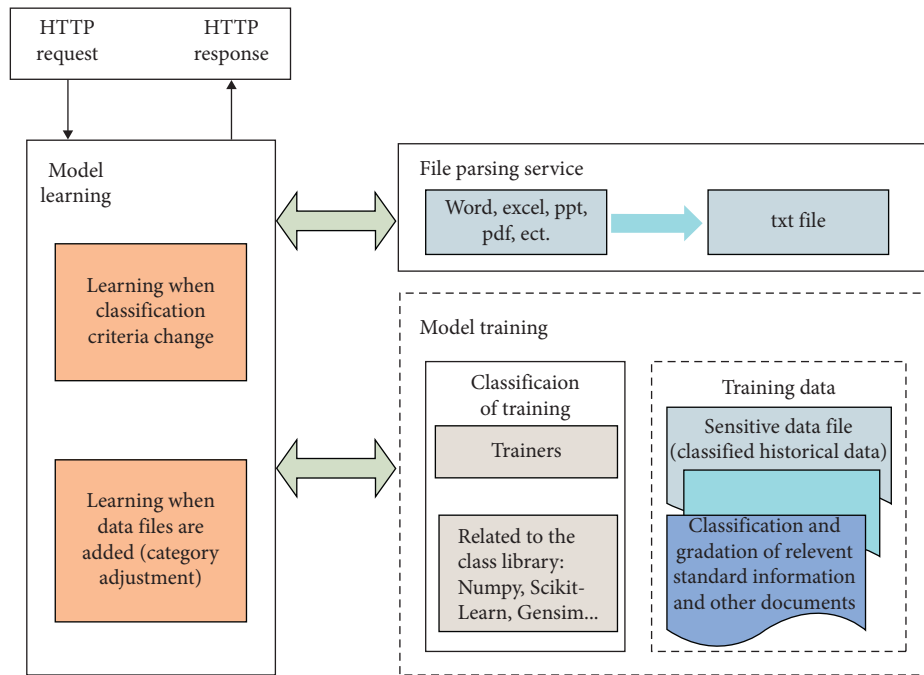


FIGURE 6: AI intelligence classification model learning frame diagram.

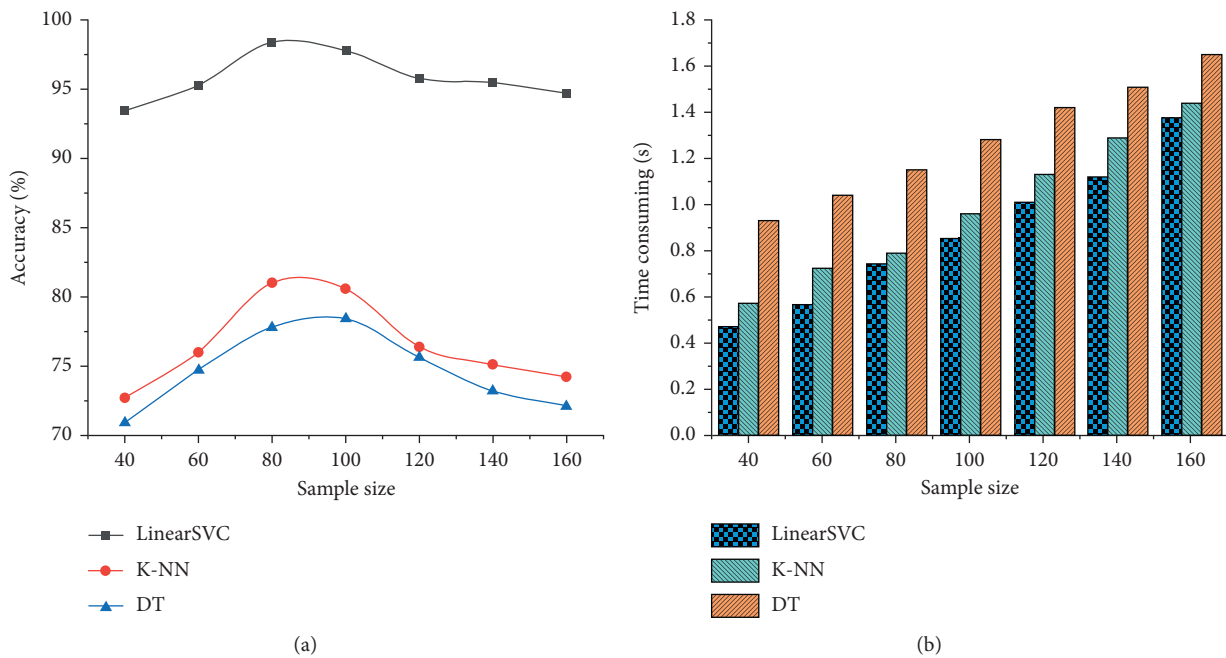


FIGURE 7: The relationship between different number of training samples and accuracy and time. (a) Accuracy line graph. (b) Modeling time diagram.

according to the nature and system of the oil company, the company roughly divides the data into 19 categories: personnel, auditing, legal affairs, and so on. At the same time, we found 19 types of documents from within the oil company, and the number of each type of documents is about 100. The various documents are divided into two parts: training set and test set. The number of the two text

sets can be flexibly set, usually set as the ratio of training set and test set is 4 : 1, but the slight increase or decrease of this ratio will not affect the accuracy of model training results.

The numbers of data files for the training and test sets of each classification category are shown in Tables 1 and 2.

The supervised model training is carried out in the enterprise classification and gradation system. When the

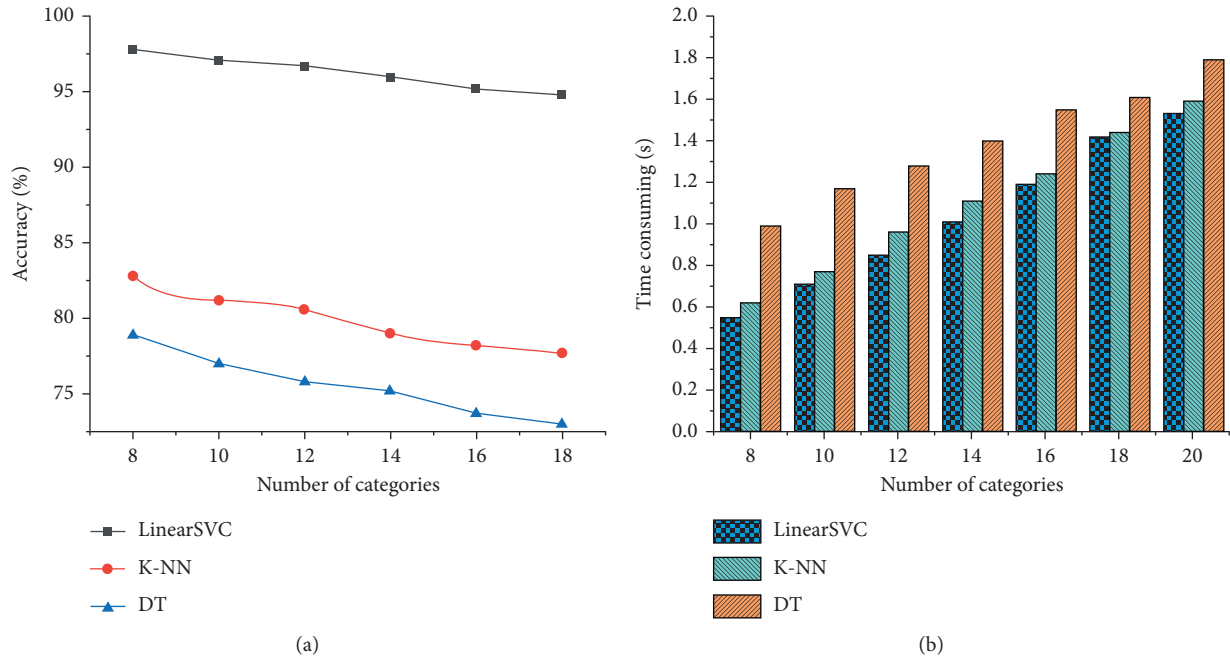


FIGURE 8: The relationship between training samples of different classification categories and accuracy and time. (a) Accuracy line graph. (b) Modeling time diagram.

TABLE 1: Number of files in each training set.

| Category                              | Quantity |
|---------------------------------------|----------|
| Professional branch                   | 71       |
| Personnel                             | 64       |
| Safety and environmental protection   | 60       |
| Global cooperation                    | 65       |
| Information management                | 62       |
| Audit                                 | 71       |
| Reform                                | 50       |
| Policy research                       | 61       |
| Legal affairs                         | 52       |
| Material procurement                  | 83       |
| Production management                 | 68       |
| Technology management                 | 99       |
| Discipline inspection and supervision | 53       |
| Letter of visit                       | 51       |
| General office                        | 49       |
| Planning                              | 65       |
| Finance                               | 76       |
| Quality and standards                 | 56       |
| Capital operation                     | 61       |
| Total                                 | 1217     |

TABLE 2: Number of files in each test set.

| Category                              | Quantity |
|---------------------------------------|----------|
| Professional branch                   | 37       |
| Personnel                             | 29       |
| Safety and environmental protection   | 21       |
| Global cooperation                    | 31       |
| Information management                | 21       |
| Audit                                 | 25       |
| Reform                                | 19       |
| Policy research                       | 23       |
| Legal affairs                         | 27       |
| Material procurement                  | 37       |
| Production management                 | 31       |
| Technology management                 | 41       |
| Discipline inspection and supervision | 33       |
| Letter of visit                       | 37       |
| General office                        | 22       |
| Planning                              | 31       |
| Finance                               | 23       |
| Quality and standards                 | 24       |
| Capital operation                     | 26       |
| Total                                 | 538      |

accuracy reaches the required level, the model is generated. Table 3 shows the accuracy and time taken for the training of the enterprise classification and gradation system model.

It can be seen from Table 3 that when the training accuracy reaches 100%, the time spent in the model training is less than 1s, and when tested, the accuracy of the model can reach 94%, exceeding 90%. Therefore, the trained model shows the advantages of short training time and high accuracy.

**5.2. Application Verification.** In order to verify the accuracy of the classification model of the system for the classification of sensitive data documents of enterprises, 3000 data files are randomly selected as test files in a petroleum enterprise. The test files were randomly divided into three groups, and the system was tested three times in succession and verified by three indicators: (3), (4), and (5).

TABLE 3: Model training results.

| Category     | Accuracy (%) | Model training time (s) |
|--------------|--------------|-------------------------|
| Training set | 100          | 0.85                    |
| Test set     | 94.8         | ×                       |

TABLE 4: Accuracy indicators of classification model.

| Test number | Precision (%) | Recall (%) | F1 (%) |
|-------------|---------------|------------|--------|
| Test 1      | 94.9          | 95.3       | 95.1   |
| Test 2      | 95.2          | 95.4       | 95.3   |
| Test 3      | 94.3          | 95.0       | 94.6   |

$$\text{precision} = \frac{\text{Num}_{\text{correct}}}{\text{Num}_{\text{total}}}, \quad (3)$$

$$\text{recall} = \frac{\text{Num}_{\text{correct}}}{\text{Num}_{\text{actual}}}, \quad (4)$$

$$F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}. \quad (5)$$

In the previous equations,  $\text{Num}_{\text{correct}}$  is the correct number of documents in the various types of data identified by the classification model,  $\text{Num}_{\text{total}}$  is the number of documents predicted by the model, and  $\text{Num}_{\text{actual}}$  is the total number of actual samples.

Table 4 clearly reveals that, in the three random testing experiments, the generated model shows great classification effect: the recall rate reaches 95% and above, and the classification accuracies are all higher than 94%. The results are consistent with the conclusions drawn in the test section, indicating that the model has good stability and can be well applied to the classification of enterprise sensitive data documents.

Regarding the part of the misclassified data that has not been verified, although the data is not classified into the correct category, it will not affect the detection of sensitive files in actual scenario. The reason is as follows: With the increase in the classification categories of sensitive documents, the difference in classification standards between the categories will gradually decrease. Therefore, in this case, it is easy to cause the misclassification of data. But, in fact, even if the data is misclassified into other categories, due to the fact that these different categories also belong to the category of sensitive documents, the purpose of preventing the leakage of sensitive documents can still be achieved.

## 6. Conclusion

This paper introduces intelligent classification technology to realize automatic classification and gradation of sensitive data of enterprises. Through the intelligent management of enterprise data, enterprises can quickly grasp the specific quantity and distribution of the information held by the enterprise, greatly reducing the learning cost of the system users and improving the efficiency of work and the accuracy of data classification and gradation. The enterprise data

classification and gradation system compensates for the problem of insufficient technical support and system adaptability of the diversified system of sensitive data classification and management through the integration of AI intelligent classification technology. But, when referring to the enterprise document classification accuracy, there is still some misclassification. Inaccurate classification will affect the accuracy of detection of sensitive documents. In the future, the accuracy of document classification should be improved in the classification and gradation system, and the misclassification should be minimized to the greatest extent to provide a more accurate pretechnical means for the enterprise data security protection and data leakage prevention.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by National Natural Science Foundation (61827811), National Defense Basic Research Program (JCKY2019407C002), the Hebei Provincial Education Departments Support Plan (SLRC2019042), the Hebei Province Funding Project for the Introduction of Overseas Students (C20200364), and China National Petroleum Corporation Information Technology Construction Project (CNPC-IT-2018-N001).

## References

- [1] J. Campos, P. Sharma, U. G. Gabiria, E. Jantunen, and D. Baglee, "A big data analytical architecture for the asset management," *Procedia CIRP*, vol. 64, pp. 369–374, 2017.
- [2] W. Song, Y. Zhang, J. Wang, H. Li, Y. Meng, and R. Cheng, "Research on characteristics and value analysis of power grid data asset," *Procedia Computer Science*, vol. 139, pp. 158–164, 2018.
- [3] Z. Liu, B. Hu, B. Huang et al., "Decision optimization of low-carbon dual-channel supply chain of auto parts based on smart city architecture," *Complexity*, vol. 2020, no. 5, 14 pages, Article ID 2145951, 2020.
- [4] T.-M. Choi, H. K. Chan, and X. Yue, "Recent development in big data analytics for business operations and risk management," *IEEE Transactions on Cybernetics*, vol. 47, no. 1, pp. 81–92, 2017.
- [5] V. V. Glukhov, I. V. Ilin, and A. B. Anisiforov, "Problems of data protection in industrial corporations enterprise architecture," in *Proceedings of the 8th International Conference on Security of Information and Networks*, pp. 34–37, Xi'an, China, 2015.
- [6] D. Core, "Applications of text classification to enterprise support documents," *Dissertations*, 2012.
- [7] S. Aier, C. Riege, and R. Winter, "Classification of enterprise architecture scenarios-an exploratory analysis," *Enterprise*

- Modelling and Information Systems Architectures (EMISA)*, vol. 3, no. 1, pp. 14–23, 2008.
- [8] K. Höne and J. H. P. Eloff, “Information security policy - what do international information security standards say?” *Computers & Security*, vol. 21, no. 5, pp. 402–409, 2002.
  - [9] S. Alneyadi, E. Sithiraseenan, and V. Muthukkumarasamy, “A semantics-aware classification approach for data leakage prevention,” *Australasian Conference on Information Security & Privacy*, vol. 8544, 2014.
  - [10] H.-C. Yan, J.-H. Zhou, and C. K. Pang, “Gaussian mixture model using semisupervised learning for probabilistic fault diagnosis under new data categories,” *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 4, pp. 723–733, 2017.
  - [11] M. Parasol, “The impact of China’s 2016 Cyber Security Law on foreign technology firms, and on China’s big data and Smart City dreams,” *Computer Law & Security Review*, vol. 34, no. 1, pp. 67–98, 2018.
  - [12] T. Lewellen, G. J. Silowash, D. L. Costa et al., *Insider Threat Control: Using Plagiarism Detection Algorithms to Prevent Data Exfiltration in Near Real Time*, Carnegie Mellon University, Pittsburgh, PA, USA, 2011.
  - [13] M. Sogodekar, S. Pandey, I. Tupkari, and A. Manekar, “Big data analytics: hadoop and tools,” in *Proceedings of the 2016 IEEE Bombay Section Symposium (IBSS)*, pp. 1–6, IEEE, Mumbai, India, 2016.
  - [14] H. Lyu, P. Li, R. Yan, H. Qian, and B. Sheng, “High-availability deployment for large enterprises,” in *Proceedings of the International Conference on Progress in Informatics & Computing (PIC)*, pp. 503–507, IEEE, Shanghai, China, 2016.
  - [15] S. Daskalaki, I. Kopanas, and N. M. Avouris, “Predictive classification with imbalanced enterprise data,” *Recent Advances in Data Mining of Enterprise Data: Algorithms and Applications*, vol. 6, pp. 147–188, 2008.
  - [16] E. Nwafor, P. Chowdhary, and A. Chandra, “A policy-driven framework for document classification and enterprise security,” in *Proceedings of the 2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress*, pp. 949–953, Toulouse, France, 2016.
  - [17] H. Liang, J. Zou, K. Zuo, and M. J. Khan, “An improved genetic algorithm optimization fuzzy controller applied to the wellhead back pressure control system,” *Mechanical Systems and Signal Processing*, vol. 142, Article ID 106708, 2020.
  - [18] H. Liang, A. Xian, M. Mao, P. Ni, and H. Wu, “A research on remote fracturing monitoring and decision-making method supporting smart city,” *Sustainable Cities And Society*, vol. 62, Article ID 102414, 2020.
  - [19] D. Ben-David, T. Domany, and A. Tarem, “Enterprise data classification using semantic web technologies,” in *Proceedings of the Semantic Web-ISWC-International Semantic Web Conference*, 2010.
  - [20] H. Liang, J. Zou, Z. Li, M. J. Khan, and Y. Lu, “Dynamic evaluation of drilling leakage risk based on fuzzy theory and PSO-SVR algorithm,” *Future Generation Computer Systems*, vol. 95, pp. 454–466, 2019.
  - [21] A. V. Savchenko, *Search Techniques in Intelligent Classification Systems*, Springer International Publishing, Berlin, Germany, 2016.
  - [22] Z. Huang, X. Xu, J. Ni, H. Zhu, and C. Wang, “Multimodal representation learning for recommendation in Internet of things,” *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10675–10685, 2019.
  - [23] A. Abu-Errub, “Arabic text classification algorithm using TFIDF and chi square measurements,” *International Journal of Computer Applications*, vol. 93, no. 6, pp. 40–45, 2014.
  - [24] Y. Zhang, R. Zhu, Z. Chen, J. Gao, and D. Xia, “Evaluating and selecting features via information theoretic lower bounds of feature inner correlations for high-dimensional data,” *European Journal of Operational Research*, 2020.
  - [25] K. A. Keating and S. Cherry, “Use and interpretation of logistic regression in habitat-selection studies,” *Journal of Wildlife Management*, vol. 68, no. 4, pp. 774–789, 2004.
  - [26] S. B. Kim, K. S. Han, H. C. Rim, and S. H. Myaeng, “Some effective techniques for naive Bayes text classification,” *IEEE Transactions on Knowledge & Data Engineering*, vol. 18, no. 11, pp. 1457–1466, 2006.
  - [27] S. Li, S. Ding, and L. Qian, “The decision tree classification and its application research in land cover,” *Remote Sensing Technology & Application*, vol. 17, no. 1, 2002.
  - [28] M. Potamias, F. Bonchi, A. Gionis, and G. Kollios, “K-nearest Neighbors in uncertain graphs,” *Proceedings of the VLDB Endowment*, vol. 3, no. 1-2, pp. 997–1008, 2010.
  - [29] J. Lehečka and J. Švec, “Improving multi-label document classification of czech news articles,” *Text, Speech, and Dialogue*, Springer International Publishing, Berlin, Germany, pp. 307–315, 2015.



## Research Article

# Mechanism of Drag Reduction in Floating Plate of Paddy Field Based on CFD

Xiaozhe Yu,<sup>1</sup> Baofeng Zhang,<sup>2</sup> and Jiahan You<sup>3</sup> 

<sup>1</sup>Inner Mongolia University for Nationalities, Tongliao 028000, China

<sup>2</sup>Yangzhou University, Yangzhou 225127, China

<sup>3</sup>Jiangsu Shipping College, Nantong 226000, China

Correspondence should be addressed to Jiahan You; 820319927@qq.com

Received 17 September 2020; Revised 17 October 2020; Accepted 26 October 2020; Published 4 December 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Xiaozhe Yu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to study drag reduction mechanism in mud parts' operation of surface machine tools for paddy field, this paper takes floating plate, the main working part of laminating mechanism, as the research object and systematically analyzes the mechanism of action of elevation angle, curved angle, penetrating angle, and local microstructure of floating plate on working resistance and local fluid flow characteristics of the laminating structure based on VOF model in Fluent. Using ship mechanics theory and fluid lubrication theory, the drag reduction mechanism under different structural parameters of the floating plate is analyzed. The results show that, compared with the ordinary floating plate, the pressure difference resistance can be reduced by increasing the elevation angle by 60°, curved angle by 20°, and mud separation angle by 20°. The increase of the concave nonsmooth bottom surface structure can reduce viscous frictional resistance, and the total working resistance after structural optimization is comparatively reduced by 48.3%, with lowered hilling height in the forward direction and improved lubrication condition of the bottom surface, forming liquid lubrication effect. This study can provide theoretical references for the optimization design of muddy soil mud parts, mud-machine interaction research, and the development of paddy field laminating mechanism.

## 1. Instruction

The paddy soil treated by field steeping forms a working layer and a hard underlayer structure, and the working layer is composed of water body and mud. The presence of mud makes the paddy field machine tools bear a large working resistance during the movement. Studies have shown that 40% of the total power consumption of rice transplanters, boat tractors, etc., is used to overcome the mud resistance in the forward process [1]. Therefore, a study on the action law of the machine tool with different surface morphology and structural parameters and paddy mud means great significance for the optimization design of paddy soil mud tools.

In recent years, scholars have carried out abundant experiments and simulation studies on drag reduction mechanism in agricultural machinery. Guo Zhijun et al. studied resistance characteristics of nine bulldozing plate models by the orthogonal test of indoor soil troughs. The

results show that, compared with the traditional curved surface structure, the parabolic bulldozing plate can reduce working resistance by 4.6%, and field mouse claw bionic curved surface structure can reduce the resistance by 16.0% [2]. Under soil moisture content of 20% and with traction resistance and soil adhesion amount as investigation indicators, Tong Jin compared 9 rib-type press rollers, finding that, compared with ordinary press roller, reasonable rib-type press roller has a visbreaking ratio up to 41.08% and a drag reduction rate up to 11.75%~39.4% [3]. Onwualu studied the effect of speed on tillage tool forces by test method and evaluated the experimental results based on three theoretical models. The results show that the second model had a more general agreement with experimental observations [4]. Ucgul et al. simulated interaction between noncohesive soil and plowing machine by EDEM. When soil particle size was 10 mm, the model could accurately predict the traction and vertical force received by the plow in a

certain speed range [5]. Using the discrete element method, JB Barr et al. simulated the effect of opener rake angle on groove profile parameters, ridge height, inclined area, trench backfill, and lateral sling silt of the sandy loam soil and compared the simulation results with the soil trough test. The relative errors were 9%, 16%, 14%, 0.8%, and 9%, respectively [6]. Woodiga and Salazar et al. installed a tail plate behind a generic tractor-trailer truck model to reduce the pressure drag. The effect of the backward-facing step height between the top surface of the trailer and the tail plate on the drag reduction is studied [7]. Skonieczny established the model of soil and farming tools by DEM and showed that excavation forces due to soil accumulation are especially sensitive to cohesion in planetary excavation with various levels of cohesion [8]. Liyan et al. established a soil particle model of paddy field using discrete element software EDEM and designed the mechanical mud parts' surface with a fish scale simulative microstructure. The motion trajectory and different distribution law of mud particles were obtained by numerical simulation, and it was calculated that the total resistance of the bionic sample is 49.84% lower than that of smooth surfaces under paddy soil conditions [9].

In summary, there are many researches on drag reduction mechanism in agricultural machinery at home and abroad, and the application methods are diverse with remarkable results achieved. However, most of them study dry field operation machinery, and there are a few researches on the drag reduction mechanism of paddy field machinery. At the same time, due to the complex fluid characteristics of paddy field mud, the research method for dry field machinery is not applicable to paddy field machinery. Based on the previous studies, this paper analyzes drag reduction mechanism of the floating plate in mud operation and provides a reference for the optimization design of related mud parts, which means great significance for reducing resistance in paddy machinery forward operation and improving the working efficiency of the unit.

## 2. Physical Model and Numerical Simulation Method

*2.1. Physical Model Establishment Method.* The actual operation on the paddy field is more complicated; it is a viscous

flow field. For this reason, the physical properties of the paddy field are first considered in the simulation process, and the mud viscosity is measured. The parameters of the mud model are shown in Table 1. The clay viscosity and geometrical characteristics of the floating plate are the most important factors in generating and affecting resistance. When the floating plate travels in the paddy mud at a certain speed, the floating plate is in contact with the mud, causing the local mud to turn, forming a vortex at the head end and the bottom of the floating plate. Moreover, the boundary layer separation occurs and the mud forms a wake flow region in progression at the tail of the floating plate under the extrusion of the floating plate, which causes a pressure change in the pressure field at the tail of the floating plate and results in a pressure difference between the head end and the tail, that is, viscous pressure resistance. Therefore, according to the working characteristics of the laminating mechanism and by virtue of the relative motion principle, the floating plate can be viewed as static, and the mud makes constant turbulence movement relative to the floating plate for numerical simulation. After the mud hits the floating plate, a turbulent flow will form in the floating plate, which is essentially a turbulent process. The mud particles affect each other, while velocity and pressure vary with time and space. Due to the severe turbulence at the head end of the floating plate, the overall flow pattern is considered as a turbulence model according to the definition of classic flow Reynolds number (Re), and wall boundary is selected in the real model. Since there is no heat transfer in the actual operation and energy form is mechanical energy, continuous equation and momentum conservation equation are selected for the governing equation, as shown in equation (1) and equation (2) [10, 11]. Eddy viscosity model is used in the numerical solution process. The standard k-model is used to solve the RANS equation, and RNG k-model with the average strain rate in reaction mainstream is introduced into equation to correct the small vortex in motion [12, 13]:

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho u)}{\partial x} + \frac{\partial(\rho v)}{\partial y} + \frac{\partial(\rho w)}{\partial z} = 0, \quad (1)$$

$$\begin{aligned} \frac{\partial(\rho u)}{\partial t} + \frac{\partial(\rho uu)}{\partial x} + \frac{\partial(\rho uv)}{\partial y} + \frac{\partial(\rho uw)}{\partial z} &= \frac{\partial}{\partial x} \left( \mu \frac{\partial u}{\partial x} \right) + \frac{\partial}{\partial y} \left( \mu \frac{\partial u}{\partial y} \right) + \frac{\partial}{\partial z} \left( \mu \frac{\partial u}{\partial z} \right) - \frac{\partial p}{\partial x} + S_u, \\ \frac{\partial(\rho v)}{\partial t} + \frac{\partial(\rho vu)}{\partial x} + \frac{\partial(\rho vv)}{\partial y} + \frac{\partial(\rho vw)}{\partial z} &= \frac{\partial}{\partial x} \left( \mu \frac{\partial v}{\partial x} \right) + \frac{\partial}{\partial y} \left( \mu \frac{\partial v}{\partial y} \right) + \frac{\partial}{\partial z} \left( \mu \frac{\partial v}{\partial z} \right) - \frac{\partial p}{\partial y} + S_v, \\ \frac{\partial(\rho w)}{\partial t} + \frac{\partial(\rho wu)}{\partial x} + \frac{\partial(\rho wv)}{\partial y} + \frac{\partial(\rho ww)}{\partial z} &= \frac{\partial}{\partial x} \left( \mu \frac{\partial w}{\partial x} \right) + \frac{\partial}{\partial y} \left( \mu \frac{\partial w}{\partial y} \right) + \frac{\partial}{\partial z} \left( \mu \frac{\partial w}{\partial z} \right) - \frac{\partial p}{\partial z} + S_w, \end{aligned} \quad (2)$$

TABLE 1: Mud model parameters.

| Moisture content (%) | Mud viscosity (Pa·s) | Mud solid phase particle size ( $\mu\text{m}$ ) | Soil density (g/ml) | Travel speed (m/s) |
|----------------------|----------------------|---|---------------------|--------------------|
| 36.1                 | 4.53                 | 100   | 2.34                | 1.44               |

where  $S_u$ ,  $S_v$ , and  $S_w$  are generalized source terms.

**2.2. Two-Phase Flow Model Establishment Method.** VOF method (Volume of Fluid) determines the free surface tracking fluid change by studying the fluid and mesh volume ratio function in the grid unit [14, 15]. In the solution, the volume fraction of a fluid in a single grid is usually kept updated without changing the interface height in interface tracking. With the advantages of no topological constraints, it is suitable for stratified flow, surface flow, and large bubble flow, thus widely used in multiphase flow models [16]. In the solution process, the volume fraction of liquid in each grid varies from 0 to 1, with 0 representing no such fluid in the grid, 1 representing the liquid filling the grid, and 0.5 representing the interface of the two liquids. The corresponding volume fraction expression is as follows [17, 18]:

$$\frac{\partial \alpha}{\partial t} + \frac{\partial}{\partial x_j} (\alpha u_j) = 0. \quad (3)$$

**2.3. Meshing Method.** The floating plate model has a complicated structure. In order to eliminate the influence of the grid itself, the same size fluid domain is used to divide the grid, but only local encryption is performed. At the same time, to deal with the interface problem of segmentation interval, workbench is adopted for direct 3D modeling [19, 20]. Wall surface partition and meshing are performed via space claim. Since the model thickness is 10 mm, according to the boundary layer characteristics, to accurately reflect the model features, the inner wall mesh is set to a minimum of 3 layers of mesh, so the grid size of the floating plate is 4 mm, the surrounding of the floating plate is filled with 15 mm grid, and area away from the stamping is set to 25 mm. The mesh generation diagram is shown in Figure 1(a). The maximum number of grids is about  $101 \times 10^4$ . The model grid computing domain is shown in Figure 1(b).

**2.4. Grid Independence Verification.** In order to avoid the influence of the size of the grid on the simulation results, four different grid division schemes were selected in turn to verify the grid independence. The model of the stamping board is established by Solidworks. Mesh preprocessing is generated by mesh. High precision grid structured grid and local encryption. The corresponding grid numbers are 851914, 1013400, 1329640, and 1720592 and the entry angle is  $50^\circ$ . Using the grid of fluid motion in the process of membrane plate in contact with soil conditions was analyzed, and the resistance change with iterative is shown. It is seen from Figure 2 when the number of meshes more than 1,013,400, the stable flow field, the calculated resistance is almost the same, the error  $< 2\%$ . Therefore, the number and

degree of encryption of the mesh calculation are the same as before.

**2.5. Boundary Condition Setting and Geometric Model Construction.** The floating plate has a movement state in the mud similar to that of the boat in the sea. The common feature of both is that small objects move in a large space. In the calculation domain and initialization model, the distance from the entrance to the model is 5 times that of board length, and the distance from the exit to the model is 3.5 times that of board length. The air zone model has a set height of 1.5 L and a lower water disturbance depth of 400 mm (equal to the board length). The immersion depth is also a value that needs special consideration, which affects the resistance value of the flow process. It can be calculated by the Baker formula, as shown in equations (4) [21, 22]. According to the field operation requirements, the immersion depth of the floating plate is generally 10~30 mm, which is affected by parameters such as mud, transplanting equipment, and degradable film weight. Therefore, the immersion depth is set to 20 mm, which can reflect the resistance value in a larger range:

$$H = \left[ \frac{G}{BL(k_c / (B + K_\phi))} \right]^{(1/n)}, \quad (4)$$

where  $H$  is the immersion depth of the floating plate,  $G$  is the weight of the floating plate (including hydraulic equipment and degradable film),  $B$  is the grounding width of the floating plate,  $L$  is the grounding length of the floating plate,  $k_c$  is the cohesive deformation modulus of the soil,  $k_\phi$  is the soil friction deformation modulus, and  $n$  is the soil deformation index.

During the simulation, the position corresponding to the immersion depth is the free interface. According to the VOF model, there is a two-phase flow interface when volume fraction is 0.5, and mud foundation depth is generally 20~40 cm when the paddy field is steeping, so the rationality and accuracy of the simulation can be guaranteed [23, 24]. The free interface height set in this paper is 400 mm and the movement speed of the floating plate is up to 6 km/h. Using a variable reference frame, the mud and air move relative to the floating plate, and the speed at the muddy water inlet and air inlet is set to 1.44 m/s. At the same time, to save the grid, both the floating plate and the air domain adopt symmetric boundaries. The remaining boundary conditions are set as follows:

- (1) The air and muddy water inlets are velocity inlets with a speed of 1.44 m/s.
- (2) The outlet is a restricted pressure outlet.
- (3) There are four wall surfaces, the rear sidewall surface, the bottom wall surface, the upper wall surface, and the wall surface of the floating plate respectively.

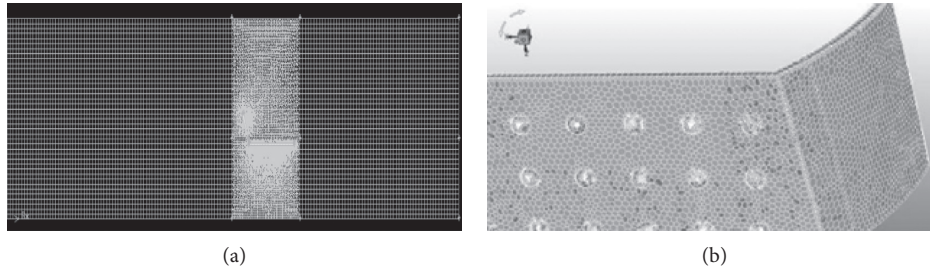


FIGURE 1: Grid division. (a) Mesh generation. (b) Convex hull local mesh refinement.

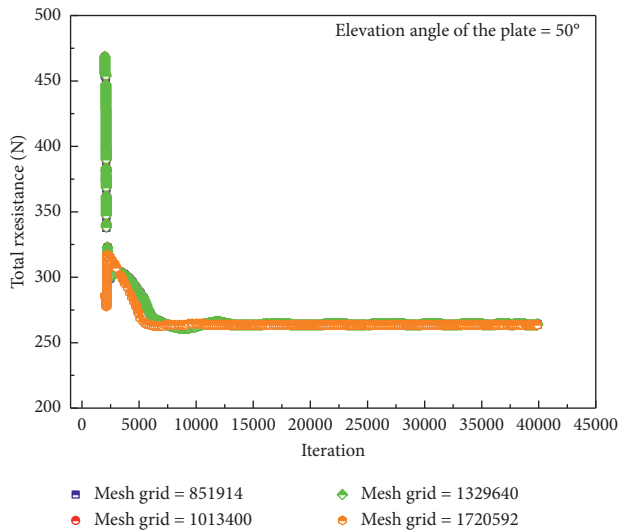


FIGURE 2: Grid independence verification.

- (4) The symmetry boundary is set in a symmetrical boundary condition for free energy and mass transfer.

### 3. Results and Discussion

Figure 3 shows a dynamic evolution diagram of the volume fraction during the movement of the floating plate. It can be seen from the figure that, with the flow of muddy water, the original static soil begins to move forward, and the soil moving in the horizontal direction hits the bottom of the floating plate. As the floating plate has a certain width and height, after blocked by the floating plate, the muddy water near the bottom of the floating plate is first decelerated, with the velocity dropped to zero. Subjected to inertia and viscosity, the subsequent muddy water cannot be stopped. Therefore, a detour convex hilling is formed at the bottom of the floating plate. Subsequent muddy water cannot continue to move in the horizontal direction, so a detour flow is formed from the bottom and the sides. The height of the hilling will continue to increase for a long time, until the hilling does not rise and stabilizes under the combined effects of gravity, viscous force, and bypass flow on both sides. Due to the influence of gravity during the ascending process of hilling, the angle of the front end boat deck affects

the rising height of the hilling. Therefore, this paper first studies the influence of the angle and height of the boat deck on the hilling height and analyzes the influencing factors affecting hilling and forward resistance.

*3.1. Influence of the Front End Angle of the Floating Plate on the Hilling.* Because of the space size limit, the floating plate is fixed; a too long length will cause inconvenience. Therefore, in the study of this paper, the floating plate front end length is selected as 160 mm with reference to analyzing the evolution law of hilling height with angle after stabilization. For a too small angle, the floating plate size will be greatly increased, and the too large volume is not conducive to operation. Therefore, the paper studies from the angle which is  $35^\circ$ . It can be seen from Figures 4 and 5 that when the boat deck angle is  $35^\circ \sim 50^\circ$ , the tilling will flow over the entire front end boat deck. After the angle reaches  $60^\circ$ , tilling is only partially piled up and will not flow over the front end boat deck. This is because, while the muddy water hits the floating plate, the splash height is affected by the angle and height of the front end boat deck of floating plate. The muddy water after the collision will continue to move forward due to inertia, which will produce a significant pressure on the surrounding boat deck.

It can also be seen from the partial pressure cloud map that the pressure here is significantly higher than that in other places, and the pressure will make the muddy water climb along the boat deck. During the climb, the muddy water decelerates and falls under the action of gravity. When the energy for the upward movement is exhausted, hilling will stop. On the one hand, as the angle increases, the component force of gravity increases along the boat deck, which will cause the climbing height to decrease. On the other hand, under the same boat deck length, for a larger angle, the vertical height is higher, the potential energy required for the climb is larger, and the initial kinetic energy is larger. Therefore, there will be a tendency that hilling height decreases with the increase of the angle under the same height of the deck boat.

However, the front end angle of the floating plate is not better as it is higher. From the total resistance variation chart (Figure 6), it can be seen that as the angle increases, the resistance increases and the floating plate travels difficultly. Hence, a lower angle should be selected for an appropriate length. It can be seen from the volume fraction cloud diagram that when the angle is less than  $60^\circ$ , the boat deck will

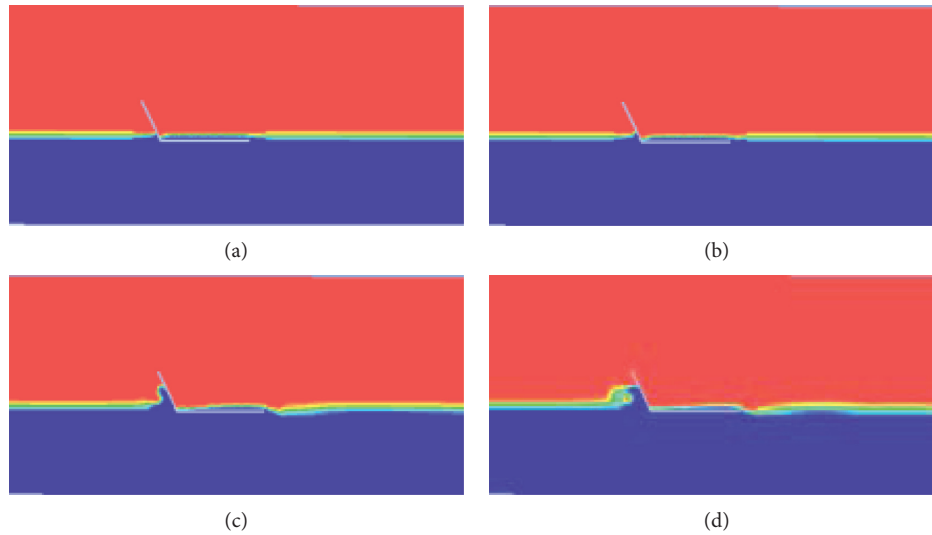


FIGURE 3: 65° floating plate simulation results: (a) 0 s, (b) 0.5 s, (c) 0.1 s, and (d) 2 s.



FIGURE 4: Simulation results of 50° floating plate. (a) 50° free waveform diagram. (b) 50° volume fraction diagram.



FIGURE 5: Simulation results of 60° floating plate. (a) 60° free waveform diagram. (b) 60° volume fraction diagram.

be submerged, and the hilling will turn over the floating plate, which is unfavorable for operation. Nevertheless, when the angle is greater than 60°, the hilling will not turn over the boat deck. Therefore, the angle should be 60° in the design and this angle is maintained for subsequent designs. It can be seen from the simulation results that the resistance of the floating plate is divided into differential pressure resistance and frictional resistance. Due to the high viscosity muddy water characteristics, the differential pressure resistance is slightly larger than the frictional resistance. It can be known from classical fluid mechanics that viscous resistance is directly related to the contact area, so reducing the contact area may be an effective factor for lower total resistance; the front end bottom of the floating plate is subjected to a large pressure, so optimization on it can reduce

the differential pressure resistance, thus overall reducing travel resistance of the floating plate and improving the working efficiency of the laminating machine.

*3.2. Influence of the Curved Angle of the Floating Plate on the Hilling Effect.* The influence of different curved angle on the resistance of the floating plate is studied by ship theory. The angle range is 5~20°. Figure 7 shows a variation chart of the floating plate resistance under different curved angles. It can be seen from the figure that the frictional resistance basically does not change with the increase of the curved angle, while differential pressure resistance is significantly reduced. Compared with an ordinary floating plate of 0°, the total resistance is reduced by about 11.90~18.18%, where the

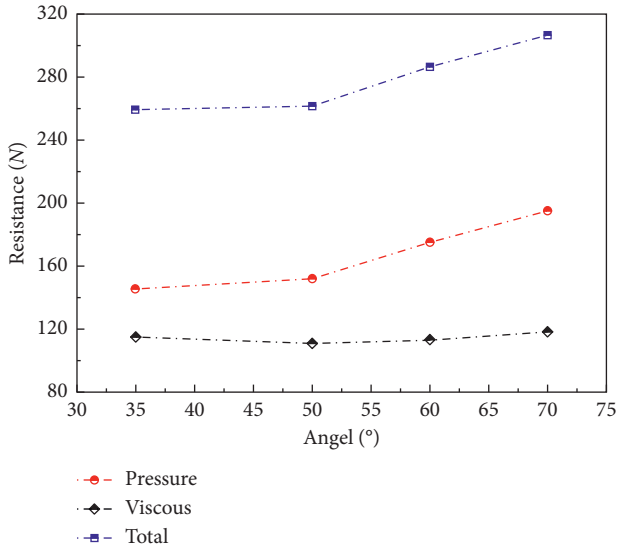


FIGURE 6: Total resistance variation chart.

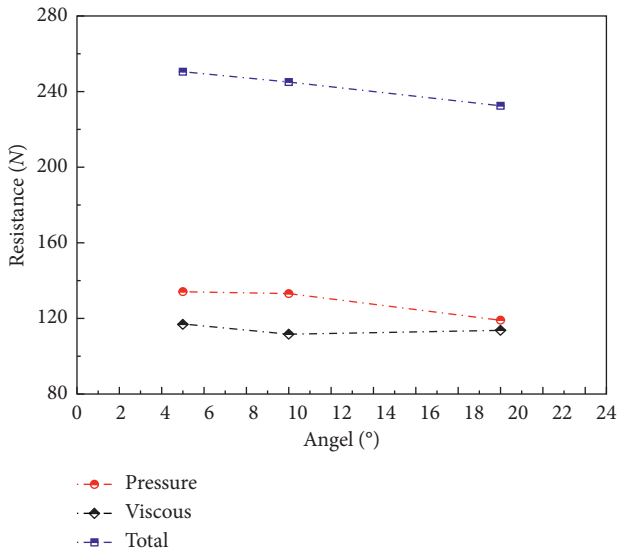


FIGURE 7: Resistance variation chart.

differential pressure resistance is significantly reduced when the curved angle is  $20^\circ$ . It can be seen from the static pressure diagram (Figure 8) that the front end boat deck of the floating plate has a big pressure drop, there is a low-pressure zone at the tail end, and the pressure difference between the two becomes pressure drop resistance. The curved surface boat deck increases the local contact area, buffers the local flow field, reduces the local pressure drop, and reduces the resistance. The pressure decrease also causes the lifting height of muddy water to drop. It can be seen from the volume fraction cloud diagram that, compared with Figure 5, the hilling height has decreased significantly, and the vortex caused by the gas-liquid two-phase separation at the edge is reduced.

It should be pointed out that, from the simulation results, it can be seen that the curved floating plate has a significant effect on the hilling effect, although a smaller

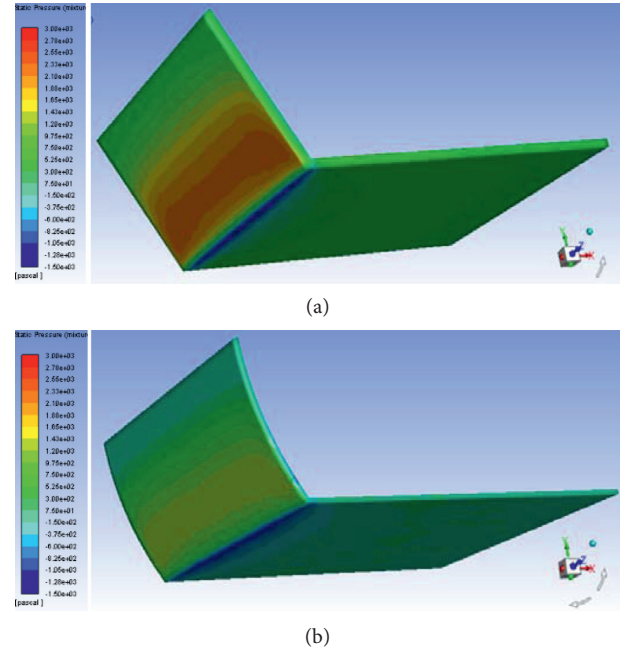


FIGURE 8: Static pressure distribution diagram: (a) static pressure distribution diagram of floating plate and (b) static pressure distribution diagram of floating plate with 20 degrees of curved angle.

angle can play a certain role and the smaller curved angle has little effect on the real situation. Therefore, it is recommended that the surface angle should be designed at around  $20^\circ$ . This is because the resistance will be increased when the angle exceeds a certain value, and at the same time, an excessively large curved surface also increases the cost and difficulty of processing. Excessive annealing will reduce the service life of the floating plate and the floating plate angle should be set to about  $20^\circ$  based on comprehensive consideration.

**3.3. Influence of the Penetrating Angle of the Floating Plate on the Hilling Effect.** In the ship dynamics, in order to reduce the resistance during the ship movement, the front end of the boat deck is usually in sharp angle to increase the momentum of the ship in the forward direction, breaking the waves and increasing the ship speed [23, 24]. In view of this, a similar penetrating angle of the floating plate structure was designed, and the penetrating angle is in the range of  $5^\circ \sim 20^\circ$  considering the paddy field working environment. Figure 9 shows the influence law of different penetrating angles on hilling height above the middle symmetry plane of the floating plate. It can be seen from the figure that hilling height is reduced with the increase of the penetrating angle, and the surface penetrating angle has a significant effect on reducing hilling height. This is because after the muddy water hits the floating plate, under the sharp interface cut, the fluid is easily split and easily flows to both sides. In this way, hilling flows around the floating plate from both sides, thus reducing the forward resistance. At the same time, in the case of a high penetrating angle, it is found that typical

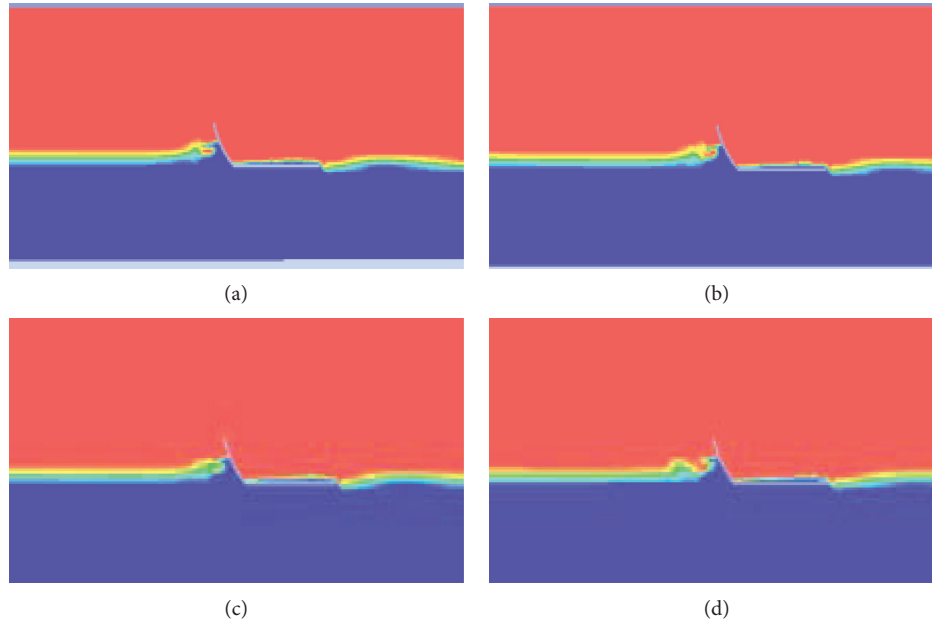


FIGURE 9: Simulation results of different penetrating angle: (a) penetrating angle is  $5^\circ$ , (b) penetrating angle is  $10^\circ$ , (c) penetrating angle is  $15^\circ$ , and (d) penetrating angle is  $20^\circ$ .

hilling is reversed back to the front, and the hilling is completely broken.

It can be seen from the local velocity cloud diagram (Figure 10) that the downward impact speed of the fluid along the floating plate is significantly higher than that of an ordinary floating plate, which also indicates that the floating plate breaks the muddy water on the one hand and accelerates fluid's downward movement on the other hand, thus reducing the hilling height. As can be seen from Figure 11, the pressure difference resistance is maximally reduced by about 21.96% as compared with the maximum of the floating plate without a penetrating angle. The calculation results also show that when the penetrating angle is  $5^\circ$ , great changes occur as compared with ordinary curved surface floating plate, with a drop of about 5.89%, and when the penetrating angle is increased, the effect becomes more and more obvious, but there is abnormal rise when the penetrating angle is 10 degrees, which may be caused by the muddy water break distance at that time.

It can be seen that the viscous resistance does not change basically. It can also be seen from Figure 12 that, with the presence of a penetrating angle, the local maximum pressure distribution changes, which changes from basically uniform distribution to a distribution with the highest static pressure at the position of the penetrating angle. This is consistent with the basic physical common sense, indicating that the result is correct. It is also necessary to point out that the tip of the penetrating angle must be sufficiently strengthened. Otherwise, there will be damage under frequent pressure impact. Therefore, it is recommended to add reinforcing structures such as ribs inside.

**3.4. Influence of Local Microstructure of Floating Plate on Resistance.** Figure 13 shows the volume fraction cloud diagram of the symmetry plane position. It can be seen from the above

figure that the surface of the three nonsmooth structures does not have a significant effect on the hilling height, but only slight difference exists. It can be seen from the total resistance distribution diagram (Figure 14) that the groove structure resistance has a decrease of 23.05 N compared with the  $20^\circ$  penetrating angle surface, and the concave structure has a decrease of nearly 45 N, while the convex groove structure not only has no resistance drop but instead shows the abnormal phenomenon of increase in resistance. The pressure drop of the three nonsmooth surfaces is basically the same in a detailed analysis, and the convex hull structure has a slight rise, but the rise is not obvious. This is because it can also be seen from the partial pressure cloud diagram of the convex hull structure that the front end facing muddy water of the convex hull structure has a pressure significantly larger than that of the tail. Under the action of a plurality of convex grooves (18), the pressure drop is significantly increased. Although the groove has small increased pressure in the impact surface facing the muddy water, the overall pressure value does not change much. At the same time, it is found that the viscous resistance of the three has a clear difference, the obvious viscous resistance drop of the concave envelope contributes to almost all the resistance drop, indicating that the drag reduction effect is obvious, the viscous resistance of the groove has a slight decline, and the downward trend is not obvious; at the same time, the pressure of the convex hull increases significantly, rising by about 5.2%. Therefore, under the joint action of the two, the resistance of the concave envelope is significantly reduced, while the resistance of the convex hull is significantly increased. The nonsmooth structure at the bottom of the surface primarily reduces the total resistance by reducing the viscous resistance.

Figure 15 is a streamline diagram of three different surfaces. It can be seen from the figure that the flow fields of the three nonsmooth surfaces are completely different. The convex hull surface will show obvious bypass flow

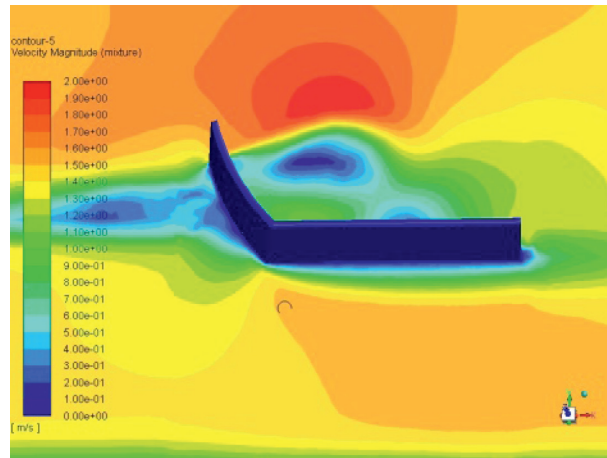


FIGURE 10: Velocity nephogram of 20.

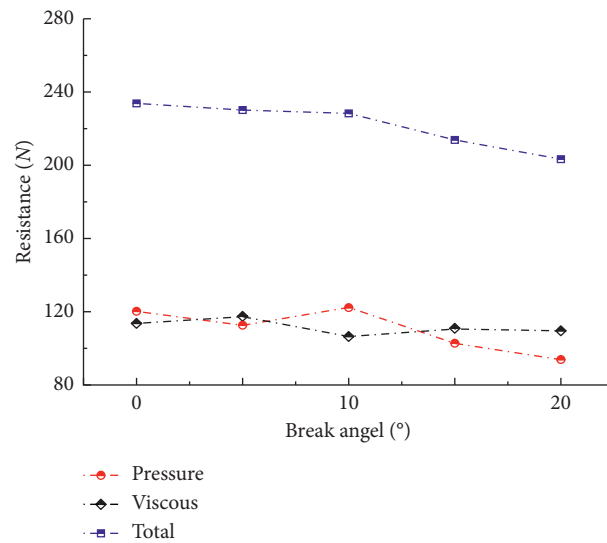


FIGURE 11: Drag distribution picture.

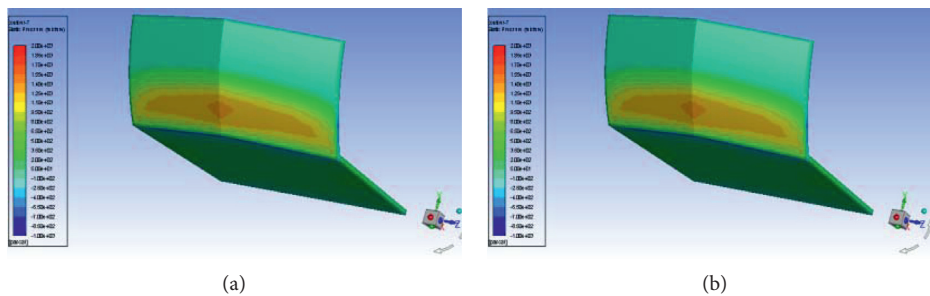


FIGURE 12: Pressure cloud diagram under different penetrating angles: (a) pressure cloud diagram when penetrating angle is 5° and (b) pressure cloud diagram when penetrating angle is 20°.

when hit by the muddy water. A wake flow is formed at the rear after the bypass. Similar to the Carnot vortex, it increases local eddy current, but since there is no local rectification structure, the superimposed disturbance of multiple vortices will inevitably increase the pressure

drop, thereby improving the running resistance. At the same time, the formation of the convex hull increases the contact area of the floating plate at the bottom. For mud flow process with great viscosity, the increase of the contact area means an increase in the viscous resistance,



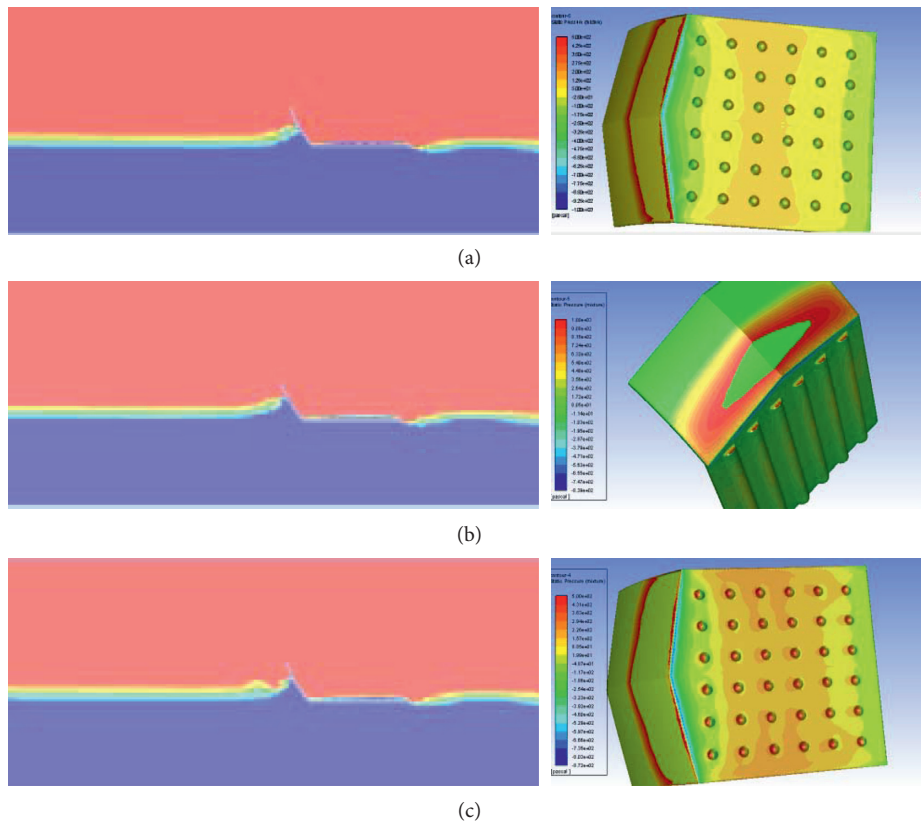


FIGURE 13: Hilling and pressure cloud diagram of nonsmooth bottom: (a) partial hilling and pressure cloud diagram of concave envelope, (b) partial hilling and pressure cloud diagram of groove, and (c) partial hilling and pressure cloud diagram of convex hull.

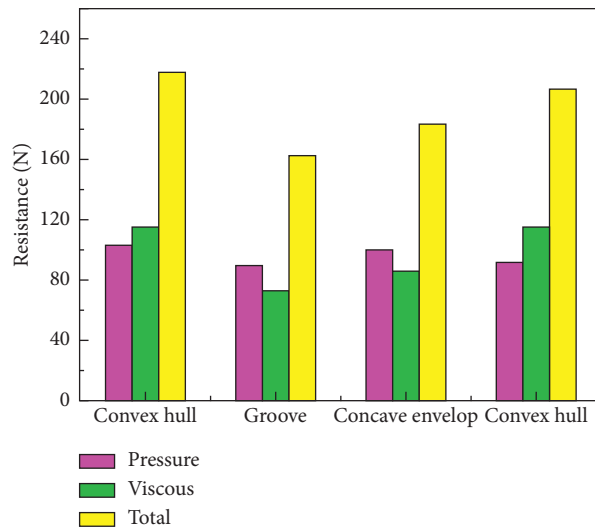


FIGURE 14: Drag distribution picture.

so the overall running resistance of the convex hull increases.

According to the simulation of the groove structure, although the contact area is increased on the surface, since the groove is long, the muddy water moves into the grooves on both sides after the impact, which accelerates the movement in the groove and reduces the viscous resistance.

For the nonsmooth structure of the concave envelope, it can be seen from the local flow field that the flow velocity near the concave envelope is extremely low, and the muddy water almost slips over the concave surface, which means that this is a typical liquid lubrication phenomenon. The muddy water filled in the concave envelope provides a lubricating effect, so that the muddy water slides down the wall surface

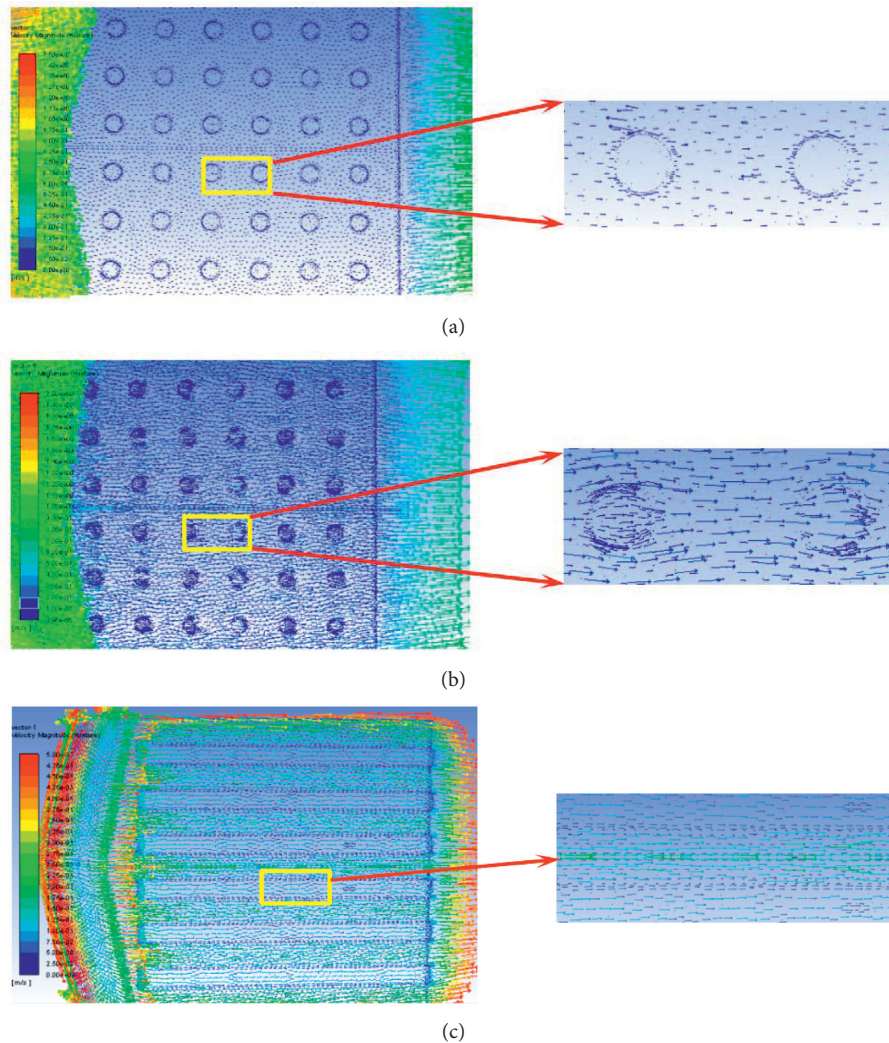


FIGURE 15: Streamline diagram of different bottom structures: (a) Velocity distribution and partial enlargement diagram of concave envelope at 365 mm. (b) Velocity distribution and partial enlargement diagram of convex hull at 365 mm. (c) Velocity distribution and partial enlargement diagram of groove at 365 mm (enlarged as groove superficial zone).

with almost no wall shearing, reducing the overall viscous resistance. Nonsmooth structures are not necessarily conducive to the reduction of total resistance but may also lead to an increase in overall resistance, such as convex hull structure. The reasonable design will reduce the resistance to a certain extent and improve the overall operating efficiency of rice laminating machine, such as concave envelope structure. In summary, we obtained the optimal design parameters for laminating machine under high viscosity in this study. The boat deck angle is  $60^\circ$ , the curved surface radian is  $20^\circ$ , the penetrating angle is  $20^\circ$ , and the bottom is in concave envelope structure.

#### 4. Experimental Verification

In order to verify the accuracy of the simulation results, this paper measures the resistance of different structures of floating plate in the indoor soil trough as Figure 16 shows. The main types of equipment for the test included

electric tractor, suspension mechanism, MIK-LCS1 tension-compression sensor, paddy soil, and floating plate model. Other types of equipment included ruler, stopwatch, marking pen, soil moisture detector, insulating tape, and assembly tools (wrench, screwdriver, plier, etc.).

During the test, the three models were moved from one end of the soil tank at a constant speed, and the moving distance was 1.5 m. Observe and measure the depth of the model sinking into the mud. After moving the model to 1.5 m, we observe the front, back, and two sides of the model, then record the measured results, and smooth the mud surface. Each model was repeated three times, and the results were averaged.

**4.1. Test Results.** Table 2 shows the traction resistance values measured after different optimization mechanism parameters. It can be seen from the table that the traction resistance



FIGURE 16: The test soil bin.

TABLE 2: Test results of traction resistance of floating plate.

| Floating plate structure type | Traction resistance (N) |              |              | Average value | Simulation result (N) |
|-------------------------------|-------------------------|--------------|--------------|---------------|-----------------------|
|                               | Test value 1            | Test value 2 | Test value 3 |               |                       |
| Common structure              | 377.3                   | 381.9        | 351.1        | 370.1         | 312.8                 |
| Curved structure              | 284.0                   | 270.1        | 262.8        | 272.3         | 236.2                 |
| Penetrating angle structure   | 240.3                   | 250.5        | 235.8        | 245.2         | 210.0                 |
| Baseplate microstructure      | 203.5                   | 195.6        | 209.6        | 202.9         | 172.7                 |

of the other three improved structures is significantly lower than that of the common structure.

On the basis of the curved structure, the traction resistance is reduced by about 27 N after the addition of the penetrating angle structure. The average traction resistance decreases to 202.9 N after adding the bottom concave envelope microstructure on the basis of the addition of a penetrating angle structure. This shows that, after several structural optimizations, there is a drag reduction effect on the working resistance of the floating plate. At the same time, the relative errors of the simulated and experimental values in the structural resistance of the four types of floating plates were 18.3%, 15.27%, 16.76%, and 17.47%, respectively. In summary, the test results are consistent with the simulation analysis results, and the simulated values are well fitted with the experimental values. Therefore, it is considered that the established interaction model between the mud parts and paddy soil based on the VOF model is reasonable.

## 5. Conclusion

In this paper, the VOF-coupled level-set model in Fluent is used to systematically study the drag reduction mechanism of the design parameters for the floating plate of the paddy field laminating machine. By investigating the influence mechanism of floating plate angle, curved angle, penetrating angle, and local nonsmooth structure on the running resistance and flow pattern of the floating plate, the research

shows that (1) the angle and height of the floating plate affect the hilling height. Although a smaller angle means small resistance, hilling will easily turn over the boat deck, reducing the effect. Comprehensively considering the hilling effect and resistance, better design angle is  $60^\circ$  under high viscosity; (2) the change of the curved surface shape reduces the pressure difference resistance. Combining theoretical model analysis values, the optimal resistance operation range can be obtained when the floating plate curved angle is  $20^\circ$ ; (3) the increase of the penetrating angle reduces hilling height, lowers running resistance, and improves the operation efficiency. Comprehensively considering local strength, the better design value of penetrating angle is  $20^\circ$ ; (4) the local nonsmooth structure has a significant influence on the viscous resistance. Comparing the three nonsmooth structures, concave envelope has the best drag reduction effect due to the local liquid lubrication. The simulation results were verified through an indoor soil trough test, which realized the innovation of the research method for the drag reduction mechanism of paddy field mud parts. The established model and the data obtained by the simulation result will contribute to the research and optimization design of the drag reduction mechanism of paddy field tools.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This study was supported by the Inner Mongolia National Natural Science Foundation of China (2019LH05002).

## References

- [1] S. Karmakar and R. L. Kushwaha, "Dynamic modeling of soil tool interaction: an overview from a fluid flow perspective," *Terramechanics*, vol. 43, no. 3, pp. 411–425, 2005.
- [2] G. Zhijun, D. Gan, Z. Li, and X. Li, "Orthogonal experiment on resistance reduction by soil engaging surfaces of bulldozer blade," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 46, no. 7, pp. 372–378, 2015.
- [3] T. Jin and Z. Qingzhu, "Reduction of soil adhesion and traction resistance of ridged bionic press rollers," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 45, no. 4, pp. 135–140, 2014.
- [4] A. Onwualu, "Draught and vertical forces obtained from dynamic soil cutting by plane tillage tools," *Soil and Tillage Research*, vol. 48, no. 4, pp. 239–253, 1998.
- [5] M. Ucgul, J. M. Fielke, and C. Saunders, "3D DEM tillage simulation: validation of a hysteretic spring (plastic) contact model for a sweep tool operating in a cohesionless soil," *Soil and Tillage Research*, vol. 144, pp. 220–227, 2014.
- [6] J. B. Barr, M. Ucgul, J. M. A. Desbiolles, and J. M. Fielke, "Simulating the effect of rake angle on narrow opener performance with the discrete element method," *Biosystems Engineering*, vol. 171, pp. 1–15, 2018.
- [7] S. Woodiga, D. M. Salazar, P. Wewengkang, J. Montefort, and T. Liu, "Skin-friction topology on tail plate for tractor-trailer truck drag reduction," *Journal of Visualization*, vol. 21, no. 6, pp. 1017–1029, 2018.
- [8] K. Skonieczny, "Modeling the effects of surcharge accumulation on terrestrial and planetary wide-blade soil–tillage tool interactions," *Soil & Tillage Research*, vol. 176, pp. 104–111, 2017.
- [9] W. Liyan, Q. Sheng, T. Q. Song et al., "A DEM analysis on drag reduction characteristics of paddy field machinery surface with bionic microarchitectures," *Journal of Shenyang Agricultural University*, vol. 48, no. 1, pp. 55–62, 2017.
- [10] S. O. Wilson, R. Bryan, R. Pablo, P. I. Monserrate, and S. J. Elizabeth, "Analysis of heat transfer between a coolant fluid and a plastic blowing matrix using the ANSYS CFD tool," in *Proceedings of the World Conference on Information Systems & Technologies*, vol. 746, pp. 280–288, Naples, Italy, May 2018.
- [11] Z. Asaf, D. Rubinstein, and I. Shmulevich, "Determination of discrete element model parameters required for soil tillage," *Soil and Tillage Research*, vol. 92, no. 1–2, pp. 227–242, 2007.
- [12] A. Ibrahim, H. Bentaher, and A. Maalej, "Soil-blade orientation effect on tillage forces determined by 3D finite element models," *Spanish Journal of Agricultural Research*, vol. 12, no. 4, pp. 941–951, 2014.
- [13] P. Busato, "A simulation model for a rice-harvesting chain," *Biosystems Engineering*, vol. 129, pp. 149–159, 2015.
- [14] S. Kawale and V. P. Chandramohan, "CFD simulation of estimating critical shear stress for cleaning flat soiled surface," *Sādhanā*, vol. 42, no. 12, pp. 2137–2145, 2017.
- [15] G. Hoormazdi, J. Küpferle, A. Röttger, W. Theisen, and K. Hackl, "A concept for the estimation of soil-tool abrasive wear using ASTM-G65 test data," *International Journal of Civil Engineering*, vol. 17, no. 1, pp. 103–111, 2019.
- [16] W. Devarrewaere, D. Foqué, B. Nicolai, D. Nuytens, and P. Verboven, "Eulerian-Lagrangian CFD modelling of pesticide dust emissions from maize planters," *Atmospheric Environment*, vol. 184, pp. 304–314, 2018.
- [17] S. E. Gant and H. Tucker, "Computational fluid dynamics (CFD) modelling of atmospheric dispersion for land-use planning around major hazards sites in great Britain," *Journal of Loss Prevention in the Process Industries*, vol. 54, pp. 340–345, 2018.
- [18] W. C. Tyson and C. J. Roy, "A higher-order error estimation framework for finite-volume CFD," *Journal of Computational Physics*, vol. 394, pp. 632–657, 2019.
- [19] Su-D. Kim, J.-Y. Yoon, and S.-J. Na, "A study on the characteristics of FSW tool shapes based on CFD analysis," *Welding in the World*, vol. 62, no. 5, pp. 915–926, 2018.
- [20] A. Janečka, J. Málek, V. Průša, and G. Tierra, "Numerical scheme for simulation of transient flows of non-Newtonian fluids characterised by a non-monotone relation between the symmetric part of the velocity gradient and the Cauchy stress tensor," *Acta Mechanica*, vol. 230, no. 3, pp. 729–747, 2019.
- [21] M. K. Moraveji, M. Sabah, A. Shahryari, and A. Ghaffarkhah, "Investigation of drill pipe rotation effect on cutting transport with aerated mud using CFD approach," *Advanced Powder Technology*, vol. 28, no. 4, pp. 1141–1153, 2017.
- [22] P. Sahoo and A. Sahoo, "Hydrodynamic studies on fluidization of red mud: CFD simulation," *Advanced Powder Technology*, vol. 25, no. 6, pp. 1699–1708, 2014.
- [23] Y. K. Demirel, O. Turan, and A. Incecik, "Predicting the effect of biofouling on ship resistance using CFD," *Applied Ocean Research*, vol. 62, pp. 100–118, 2017.
- [24] M. Haase, K. Zurcher, G. Davidson, J. R. Binns, G. Thomas, and N. Bose, "Novel CFD-based full-scale resistance prediction for large medium-speed catamarans," *Ocean Engineering*, vol. 111, pp. 198–208, 2016.

## Research Article

# A New Method for Optimizing the Cabin Layout of Manned Submersibles

Wenzhong Wang <sup>1,2</sup>, Shusheng Zhang,<sup>1</sup> Cong Ye,<sup>3</sup> Dengkai Chen <sup>1</sup> and Hao Fan <sup>1</sup>

<sup>1</sup>Key Laboratory of Industrial Design and Ergonomics, Ministry of Industry and Information Technology, Northwestern Polytechnical University, Xi'an 710072, China

<sup>2</sup>Shaanxi University of Science and Technology, Xi'an, Shaanxi 710021, China

<sup>3</sup>China Ship Scientific Research Center, Wuxi, Jiangsu 214082, China

Correspondence should be addressed to Dengkai Chen; [chendengkai@nwpu.edu.cn](mailto:chendengkai@nwpu.edu.cn)

Received 20 October 2020; Revised 17 November 2020; Accepted 19 November 2020; Published 4 December 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Wenzhong Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to reduce the misoperation of submersible pilots in the complex environment of deep sea and improve human reliability, it is a very important method to optimize the cabin space of manned submersible. In this paper, manned submersible Jiaolong is taken as an example to describe the breakthrough of traditional modes for optimizing the cabin layout of manned submersible with an aesthetic perspective of deconstruction and reconstruction thinking; the layout is optimized based on deconstruction and reconstruction from the cultural perspective of interdisciplinary. The experimental results show that the layout of manned submersible cabin is optimized by combining the theory of deconstruction and reconstruction aesthetics and human factors engineering, and the feasibility and effectiveness of this method are verified by multiobjective genetic algorithm. Deconstruction and reconfiguration layout optimization improves the human reliability of divers. It is a new method for the optimization of manned submersible cabin layout and also provides a reference for studies on the layout of similar small space.

## 1. Introduction

Since the 21st century, human beings have entered a new era of ocean exploration. Currently, the “blue territory” with deep-sea resources and military strategic significance is being developed, utilized, and protected by numerous coastal countries. The first choice for ocean exploration is manned submersible, which is called “an important cornerstone of oceanographic research.” Manned submersible refers to a type of diving device with the capacities of underwater observation and operation, which can be mainly used to carry out the tasks such as deep-sea investigation, exploration, development, and salvage, and is usually taken as a base for underwater activities by submersible pilots. Deep-sea manned submersible, as the forefront and commanding height of ocean exploitation, will reflect the comprehensive strength of a country [1]. However, in the complex deep-sea environment over thousands of meters

deep, several submersible pilots shall perform the high-intensity and high-precision scientific investigation for more than ten hours in a manned submersible cabin with an inner diameter of about only 2 meters, which is almost an impossible mission. Therefore, the matter of how to optimize the space layout of a manned submersible cabin and improve the human reliability of submersible pilots has always been a hot topic in maritime powers [2].

The manned submersible cabin layout has been performing better and better through continuous upgrading and transformation by means of function orientation, ergonomics, space segmentation, and algorithm based on the preliminary design [3]. Foreign and domestic manned submersibles, such as Ross/Consul and Peace I/II of Russia, Alvin and Nadir/Deep Rover of the United States, Deep Sea 6500 of Japan, Nautilie of France and Jiaolong, Deep Sea Warrior, and Fendouzhe of China, as shown in Table 1, have been transformed as planned in recent years; they have

TABLE 1: Main technical parameters of domestic and foreign manned submersibles.

| Item          | Country           | Depth (m) | Length (m) | Height (m) | Width (m) | Total weight (t) | Manned cabin material | Cabin inner diameter (m) |
|---------------|-------------------|-----------|------------|------------|-----------|------------------|-----------------------|--------------------------|
| Peace I/II    | Russia            | 6000      | 7.8        | 3.0        | 3.6       | 18.6             | Nickel steel          | 2.1                      |
| Ross/Consul   | Russia            | 6000      | 8.0        | 3.7        | 3.7       | 24               | Ti alloy              | 2.1                      |
| Nautile       | France            | 6000      | 8.0        | 3.8        | 2.7       | 19.5             | Ti alloy              | 2.1                      |
| Deep Sea 6500 | Japan             | 6500      | 9.5        | 3.2        | 2.7       | 25.8             | Ti alloy              | 2.1                      |
| Alvin         | The United States | 4500      | 7.1        | 3.7        | 2.6       | 17               | Ti alloy              | 1.98                     |
| Jiaolong      | China             | 7000      | 8.2        | 3.4        | 3.0       | 23.5             | Ti alloy              | 2.1                      |

accepted continuous and planned upgrades in aspects of cabin layout, and all have achieved great success [4].

The optimized layout of manned submersibles, aerospace, and aircraft enclosed cabins usually takes the operator as the center to build simulators to optimize the layout. For example, Boeing’s BOEMAN, NASA’s A3I program, and NATO’s AGARD Aircraft Cockpit program all use simulators to study how to optimize space layout and reduce the misoperation caused by human factors. The space layout of manned submersible cabin adopts the traditional cut-and-trial manner; later, they turned to adopt 3D modeling with computer software for layout cut and trial, which was not materially different from a model and could not reach a satisfactory result either. The layout based on ergonomics, such as muscle fatigue of submersible pilots, would also cause a series of problems; therefore, it is a trend to optimizing cabin layout with multiple technologies oriented by manned submersible functions and ergonomics. With the increase of diving depth of manned submersibles, the cabin space will be smaller and smaller, the interference between equipment and equipment becomes greater and greater, the working hours of submersible pilots become longer and longer, and the multidisciplinary optimization design system will be highlighted during the research of cabin layout. In this system, visual simulation, virtual reality technology, and mathematical algorithm model will be introduced [5].

The optimization of manned submersible cabin layout has been continuously explored to seek new design methods and means to optimize the layout, thus improving human reliability. In this paper, the manned submersible cabin layout is optimized based on deconstruction and reconstruction thinking from the cross-disciplinary cultural perspective and then verified in combination with multi-objective genetic algorithm. It will provide a new method for the research on cabin layout of manned submersibles and also provide a reference for the research on the layout of small cabins of aircraft, ships, and space ships [6].

## 2. Application of Deconstruction and Reconstruction Thinking in Optimization of Cabin Layout of Manned Submersible Jiaolong

*2.1. Deconstruction and Reconstruction Thinking.* Deconstruction (structural decomposition) is a term proposed by Derrida, a French poststructuralist philosopher,

which was derived from “deconstruction” in Sein Und Zeit. Reconstruction refers to the rationalization of design pattern and architecture through adjusting and improving quality and performance, thus improving the scalability and maintainability. In short, deconstruction is the decomposition of the original structure into basic units, while reconstruction is the restoration of such basic units to parts for recombination, thus obtaining a brand new and different structure.

Deconstruction and reconstruction thinking is a way of thinking for getting new objects by recombination in a new way based on representative elements taken from the basic units decomposed from a specific object, such as overall appearance, local morphology, or microfeatures. Basic deconstruction methods include planarization, geometrization, symbolization, and abstraction. Reconstruction has the following types and methods: reconstruction of the concrete form: reality and superreality; and reconstruction of the abstract form: geometric abstraction (cool abstraction) and free abstraction (hot abstraction).

*2.2. Symbolic Deconstruction of Static Space in the Cabin of Manned Submersible Jiaolong.* The cabin of the manned submersible Jiaolong is approximately an oval sphere with an inner diameter of 2.1 m, which can accommodate three people, including an operator and two scientists, as well as the corresponding seats, monitors, manipulators, batteries, and the related electronics. First, qualify the ergonomic parameters of people and facilities, and evolve the objects in the spherical cabin into simple symbols based on the tasks to be performed, as shown in Table 2; later, take these symbols as the constituent elements of space layout, which is the symbolic expression of deconstruction and reconstruction thinking, as shown in Figure 1; three submersible pilots (⊙round) are arranged in the middle of the cabin, with the grey dotted circle and rectangular boxes as their range of activity; workbenches (∟lines), monitors, and maneuvering devices (□square) are arranged around, for taking full advantage of the spherical cabin wall.

As shown in Figure 1, the objects in the cabin are symbolically deconstructed, and the displays and monitors are combined with the spherical wall. As shown in Figure 2(a), the operating table and monitoring equipment shall be reconstructed in combination with Figure 2(b) by making full use of spatial features of the spherical wall

TABLE 2: Objects in the spherical cabin and occupied area.

| Name                        | Parameters                                  | Quantity | Occupied area A (m <sup>2</sup> ) | Area          | Symbolization |
|-----------------------------|---|----------|-----------------------------------|---------------|---------------|
| Submersible pilots          | 50% sample population                       | 3        | 0.25 (50% sample)                 | Activity area | ⊙round        |
| Seats                       | 600 × 400 × 350 (mm)                        | 3        | 0.252                             | Working area  | □square       |
| Monitors and other devices  | /   | Several  | /                                 | Cabin wall    | □square       |
| Console                     | /   | 1        | /                                 | Wall edge     | □square       |
| Batteries                   | 500 × 400 × 1200 (mm)                       | 1 set    | 0.24                              | Cabin bottom  | □square       |
| Storage and food containers | 20 L for normal use; 50 L for emergency use | 1 each   | 0.02; 0.05                        | Cabin bottom  | □square       |
| Life support area           | 500 × 400 × 1200 (mm)                       | 1        | 0.24                              | Cabin top     | □square       |

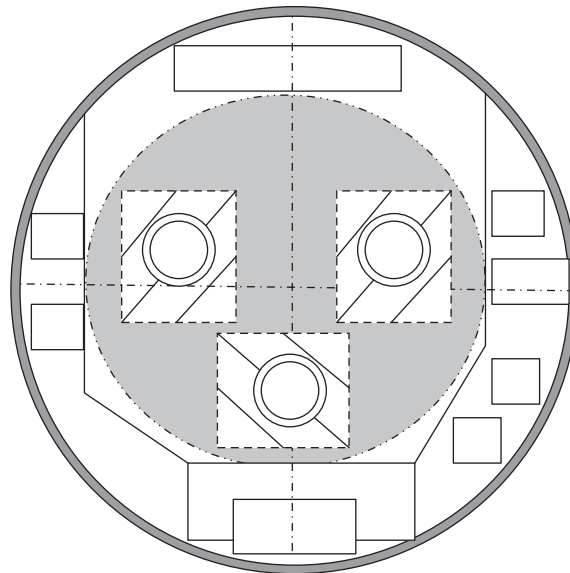


FIGURE 1: Symbolic expression in the spherical cabin.

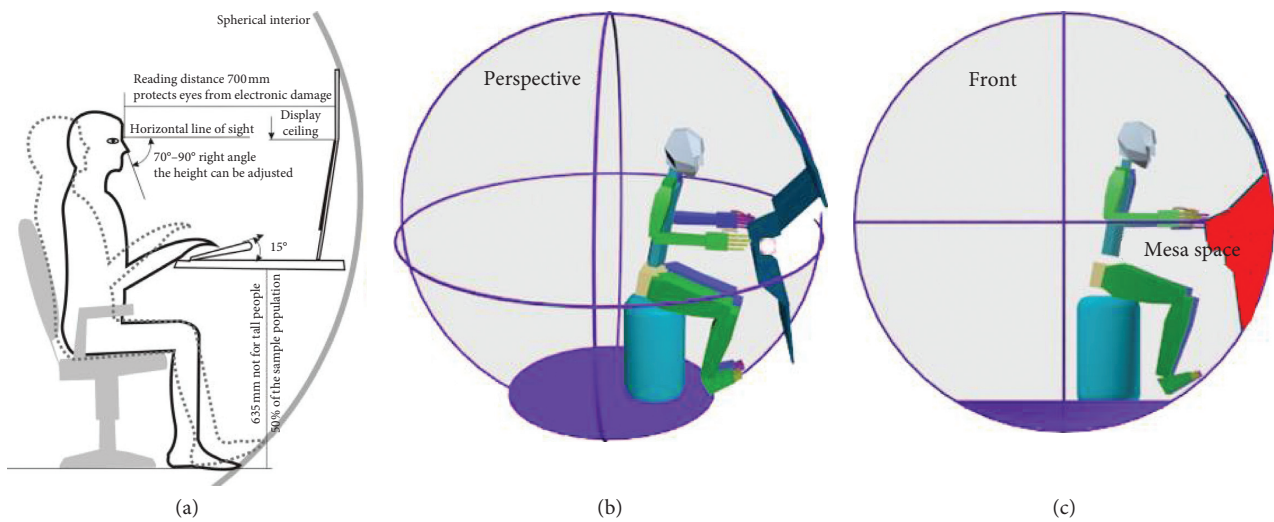


FIGURE 2: Reconstruction and optimization of the working table and display system of manned submersible cabin.

according to the ergonomic requirements for submersible pilots under the conditions of meeting the strong constraint requirements of visibility, accessibility, and comfort and then form the red space as shown in Figure 2(c), to place electronic devices and personal belongings of submersible pilots.

*2.3. Geometric Deconstruction of Active Regions in the Cabin of Manned Submersible Jiaolong.* The manned submersible cabin consists of the master control area, display area, life support area, working area, and area of other facilities. Its spatial relation can be decomposed into several geometric planes, which may be taken as basic modeling elements for performing geometric deconstruction combining with the tasks of manned submersible, the underwater environment, and working area in the cabin [7].

The layout of objects in the cabin can be divided into three parts, and most destructive collisions would be concentrated in the front and lower parts of manned submersible, so it is suggested placing energy and electronic devices in the lower part of manned submersible and back area of the cabin with great structural strength. In addition, according to the tumbler principle, the batteries shall be placed at the bottom for easy maintenance and maintain the balance of manned submersible. Due to few collisions on the top of the cabin, the upper part is relatively safe, so it can be taken as the last line of defense for the safety of submersible pilots and the place for storing life support materials, as shown in Figure 3(a). The active region mainly refers to the working area of the three submersible pilots in the middle of the spherical cabin, as shown in Figure 3(b). Through geometric deconstruction and reconstruction for submersible pilots and objects in the cabin, the display and monitoring devices shall be fixed on the spherical inner wall, so as to ensure the visibility and the operability within the control range of upper limbs, thus ensuring the accessibility of control [8].

The seats are usually reconstructed in a similar manner, which may be set as the integrated or folded geometry with geometric construction surface as the basic modeling element. The integrated seat can increase the storage space, while the folded one can increase the activity space after being folded and allow 3 people to seat after being unfolded. In this case, the integrated seat is adopted. Through graphic reconstruction of the storage space shown in Figure 3 in combination with the seats for 3 submersible pilots, the lower storage space is increased as much as possible while meeting the seating requirements of submersible pilots, as shown in Figure 4. While designing the integrated seat, the man-machine ergonomic parameters such as visibility and comfort of submersible pilots, the accessibility, and comfort of manned submersible equipment shall be constrained in strict accordance with the ergonomic requirements, so as to construct the model of the integrated seat and working table for submersible pilots as shown in Figure 4.

*2.4. Reconstruction and Optimization of Overlapping Area in the Cabin of Manned Submersible Jiaolong.* The operation of manned submersible can be simplified into five stages:

diving, sitting on the sea bed, cruising, operating, and ascending, which involve the tasks of photo taking, camera shooting, surveying, and sampling [9]. Surveying, sampling, and manned submersible manipulation would basically overlap in pairs; therefore, we should freely and abstractly reconstruct the overlapping work area and arrange movement areas for submersible pilots undertaking the overlapping tasks, thus meeting the requirements for simultaneous working by submersible pilots without interference, as shown in Figure 5. Take sampling; for example, reconstruction and optimization of overlapped areas can save the activity space for one submersible pilot and ensure that the operation and observation space are sufficient for two operators without mutual interference, and another scientist may only need to stand in front of the monitor and direct the operations while monitoring.

The working areas of the three submersible pilots are overlapped, with the activity space in the front, as shown in Figure 5. Their activities are concentrated in the front, one may perform operations, the second one may provide observation and assistance, and the third one may monitor the working status. The weight of space overlapping is the lowest at the wall and the rear half, as shown in Table 3. After deconstruction, the rear part of the cabin is used to place equipment, and through optimization and reconstruction of the integrated seat, the seat and rear space can be designed in a model conforming to ergonomics. The lower part in the rear is provided with a certain space for a submersible pilot to lie down and relax (for they will work for more than ten hours in the cabin with a diameter of only 2.1 m). The upper part of the spherical cabin is separated, the middle part is vacated, and the top is set with a passageway, thus obtaining the space layout as shown in Figure 6.

### **3. Establish the Optimization Model of Cabin Layout of Manned Submersible Jiaolong under the Deconstruction and Reconstruction Thinking**

In this paper, we consider the basic elements of the reconstruction as modules and seek the sources of manned submersible cabin module, the function of each part of the cabin represented by the module, and the form of each module in the cabin by the deconstruction and reconstruction thinking, so as to reconstruct the space layout of the cabin module system under different thinking [10].

The abstract module system is a logical system with coherent design, so there is a close relationship between task decomposition modules, and the matching of different modules may have different meanings. The module elements applying deconstruction thinking are qualified (sample population is selected for submersible pilots), as shown in Figure 2; as for cabin layout, the space of each qualified module has an exact range, and the modules can be freely replaced under the condition that the occupied space does not exceed the maximum range of activity. The overlapping weight grade is set between modules, as shown in Table 3, to further optimize the layout of cabin space.



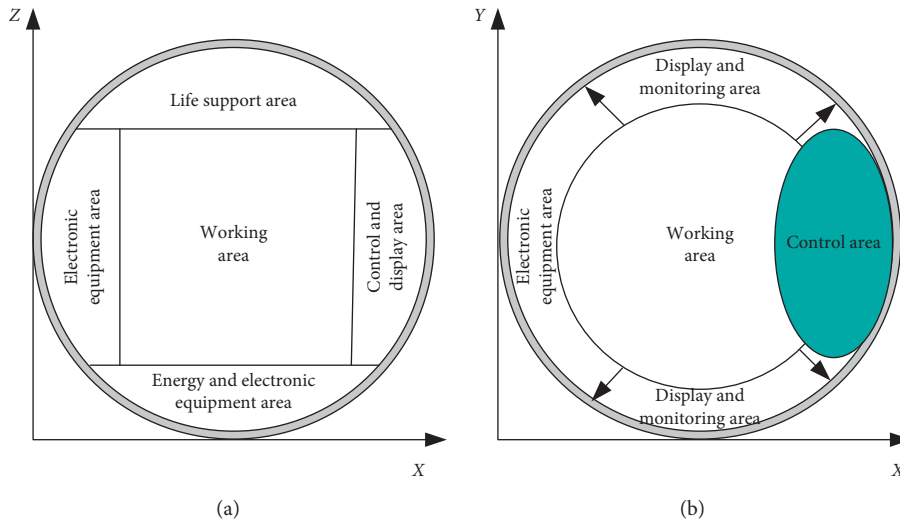


FIGURE 3: Geometric deconstruction of space in manned submersible cabin.

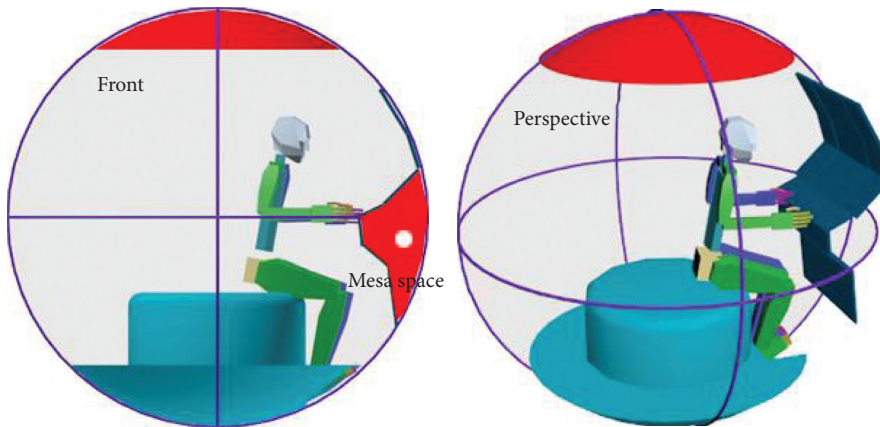


FIGURE 4: Reconstruction and optimization of the seat for three submersible pilots.

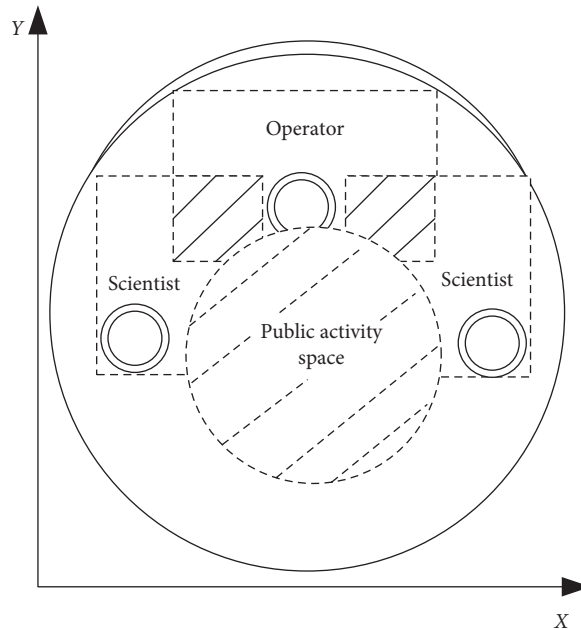


FIGURE 5: Reconstruction optimization of overlapping working areas in manned submersible cabin.

TABLE 3: Overlapping weight grade and definition of coefficients.

| Grade | Weight | Meaning   |
|-------|--------|---|
| 1     | 1      | Regional overlap demand is quite high or the regions must be overlapped |
| 2     | 0.7    | Overlap demand is relatively high                                       |
| 3     | 0.4    | Overlap demand is general   |
| 4     | 0.1    | Overlap demand is relatively low  |

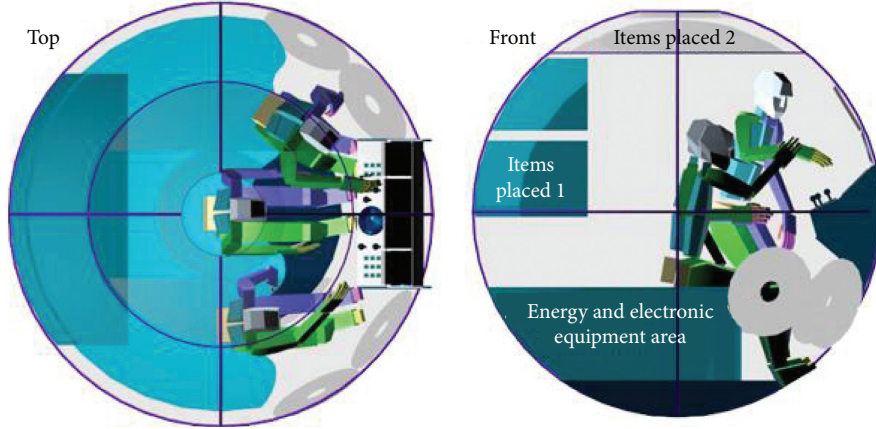


FIGURE 6: Abstract reconstruction and optimization of overlapped space for submersible pilots in the spherical cabin.

**3.1. Cabin Layout Optimization Model of Manned Submersible Jiaolong.** Suppose that  $m$  target objects (including 3 submersible pilots) shall be placed in the spherical cabin; the design variable of  $m$  objects is  $X = (x_1, x_2, \dots, x_m)^T$ , representing the spatial position of each object;  $F$  refers to the objective function, representing the design requirements of each object;  $r$  refers to the radius of the spherical cabin;  $v_i$  refers to the space occupied by objects in the spherical cabin;  $v_j$  refers to the space occupied by objects during movement;  $i$  refers to the variable of objects occupying space in the spherical cabin;  $j$  refers to the variable of objects with mutual interference of active space; and  $k$  refers to the equipment overlapping weight coefficient, as shown in Table 3:

$$\begin{cases} \max F = \{f_1(X), f_2(X), \dots, f_m(X)\}, \\ \text{Vol}_F = \sum_{i=0}^m V_i + \sum_{j=0}^m kV_j, \\ \text{s.t. } 0 \leq i, j \leq m, \\ \text{Vol}_F \leq \frac{4}{3}\pi r^2. \end{cases} \quad (1)$$

According to the ergonomic quantitative data of the cabin model, we perform preliminary deconstruction and reconstruction of the targets in the cabin and get the model of cabin layout (Figure 7) according to the requirements of noncollision between objects in manned submersible cabin, between equipment and equipment, and between equipment and cabin wall.

**3.2. Optimization of Cabin Layout under Noninterference Constraint.** Noninterference refers to the zero probability of collision between equipment and equipment in the cabin and between the equipment and cabin wall, which is the first condition for the optimization of cabin layout [11]. According to calculation formula (2), the metrics can be used to measure noninterfering collisions, where  $f_I$  refers to the total interference of space layout in the cabin,  $I_1(x)$  refers to the interference amount between the equipment and cabin wall, and  $I_2(x)$  refers to the interference amount between equipment and equipment. In summary, the mathematical model can be expressed as follows:

$$\begin{cases} \max F = \{f_1(X), f_2(X), \dots, f_m(X)\}, \\ f_I = \sum I_1(x) + \sum I_2(x), \\ \text{s.t. } |X_i| \leq R, \quad i = 1, 2, \dots, m, \\ f_I = 0. \end{cases} \quad (2)$$

#### 4. Pareto Multiobjective Optimization Genetic Algorithm

Optimization refers to getting better items through certain rules or algorithms. In a given area, the optimization of multiple numerical targets is called the multiobjective optimization. In this paper, the genetic algorithm, combined with the parallel selection strategy, is mainly used to solve the multiobjective optimization in the cabin of manned submersible Jiaolong after deconstruction and reconstruction [12].

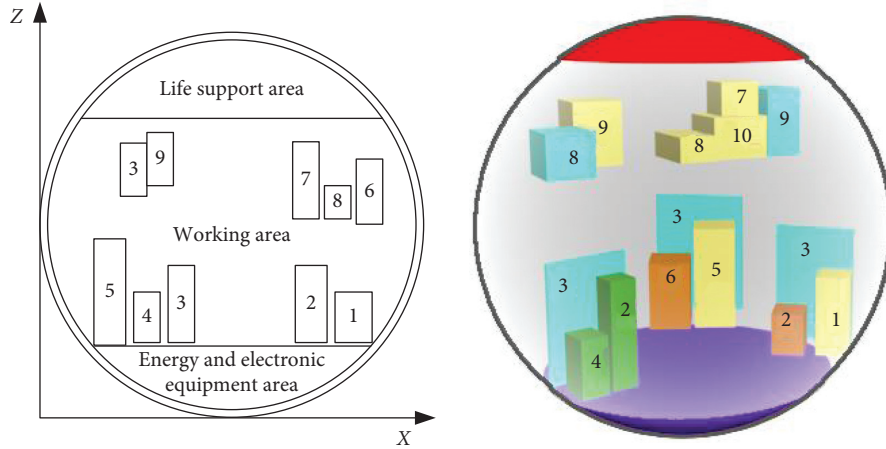


FIGURE 7: Model of space layout of equipment in the cabin.

**4.1. Improved Genetic Algorithm.** The manned submersible cabin group (all kinds of equipment) is divided into different subgroups based on the number of objective functions of cabin layout category after deconstruction and reconstruction, and the corresponding objective function is provided for each subgroup; later, the optimal subfunctions are independently selected under the constraint conditions of cabin layout and formed into the corresponding new subgroups, thus combining into new groups; and then the cross-mutation operation is performed; finally, the group of the next generation is generated, and the Pareto optimal solution will be obtained after cycling [13], as shown in Figure 8. The flowchart of parallel selection genetic algorithm after deconstruction and reconstruction is shown in Figure 9.

After deconstruction and reconstruction, the genetic algorithm can be improved from the following aspects: maintain the Pareto optimal items in the subgroups of the next generation, without participating in crossover calculation and mutation calculation. The area to be searched, such as the integrated seat and the rear item placement area after deconstruction and reconstruction, can be taken as the final search result, as shown in Figure 10; the search trajectory of Figure 10(a) improved genetic algorithm is more effective than that of Figure 10(b) random search, based on which, the fitness function value is established for guiding the fast and stable search.

**4.2. Coding and Fitness Function.** The solution variable is the coordinate of each of the types of equipment in the cabin of manned submersible Jiaolong, which can be taken as the chromosome for coding.  $X$  refers to a certain layout scheme.  $x_{ic}$ ,  $y_{ic}$ , and  $z_{ic}$  refer to the center-of-mass coordinate of each equipment:

$$X = (x_{1c}, y_{1c}, z_{1c}, x_{2c}, y_{2c}, z_{2c}, \dots, x_{mc}, y_{mc}, z_{mc}). \quad (3)$$

Fitness function value is usually used to evaluate the individual performance and guide the search in genetic algorithm, and it directly determines the stability and speed of the genetic algorithm. Take the space layout in the cabin of

manned submersible Jiaolong as an example;  $R$  of the spherical cabin is 2,000 mm, so in order to find an optimal layout scheme of the cabin of manned submersible Jiaolong, the first condition is that the devices cannot collide with each other (i.e.,  $f_I = 0$ ). The fitness function is

$$F(p) = k \times \sum_{i=1}^m Q_i \times f_i(p), \quad (4)$$

where  $P$  refers to the layout scheme of manned submersible cabin;  $Q_i$  refers to the prior layout index of the cabin after deconstruction and reconstruction;  $f_i(p)$  refers to the optimal individual value of the equipment in manned submersible layout scheme;  $k$  refers to the equipment overlapping weight coefficient.

**4.3. Main Parameters of the Optimization Algorithm.** The manned submersible group size  $N=50$ . According to formula (3),  $n$  is 10; crossover probability  $P_c=0.8$ ; mutation probability  $P_m=0.1$ ; generation gap of subgroup  $G_{Gap}=0.8$  (each generation has 20% of outstanding individuals inherited into the next generation); the maximum number of evolution generation  $G_{max}=500$ . The initial individuals are randomly arranged and automatically optimized during the operation of the algorithm. The objective function values of the subgroups are all subject to dimensionless and normalization processing [14].

**4.4. Analysis of Case Results.** In this case, the results of 20 calculations are basically consistent, verifying that the optimized layout of manned submersible cabins constructed after deconstruction and reconstruction has a better stability. A set of calculation results (only part of the results) is extracted for analysis, as shown in Figure 11.

As for Figure 11(a), optimization of equipment with low overlapping weight in the cabin, the initial objects are randomly arranged, the movement trajectory of equipment in the cabin is chaotic, and there are many collisions in the activity space, which are mainly the collisions between equipment and inner wall of the spherical cabin and between

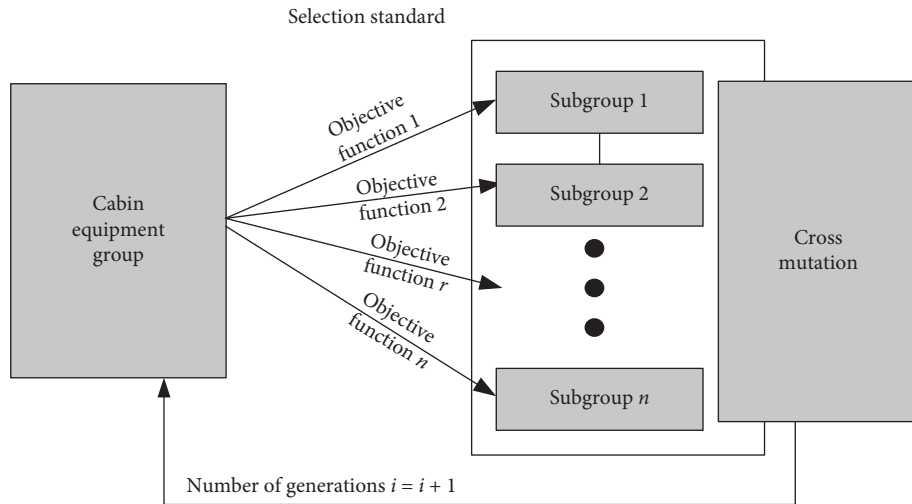


FIGURE 8: Diagram of parallel selection genetic algorithm after deconstruction and reconstruction.

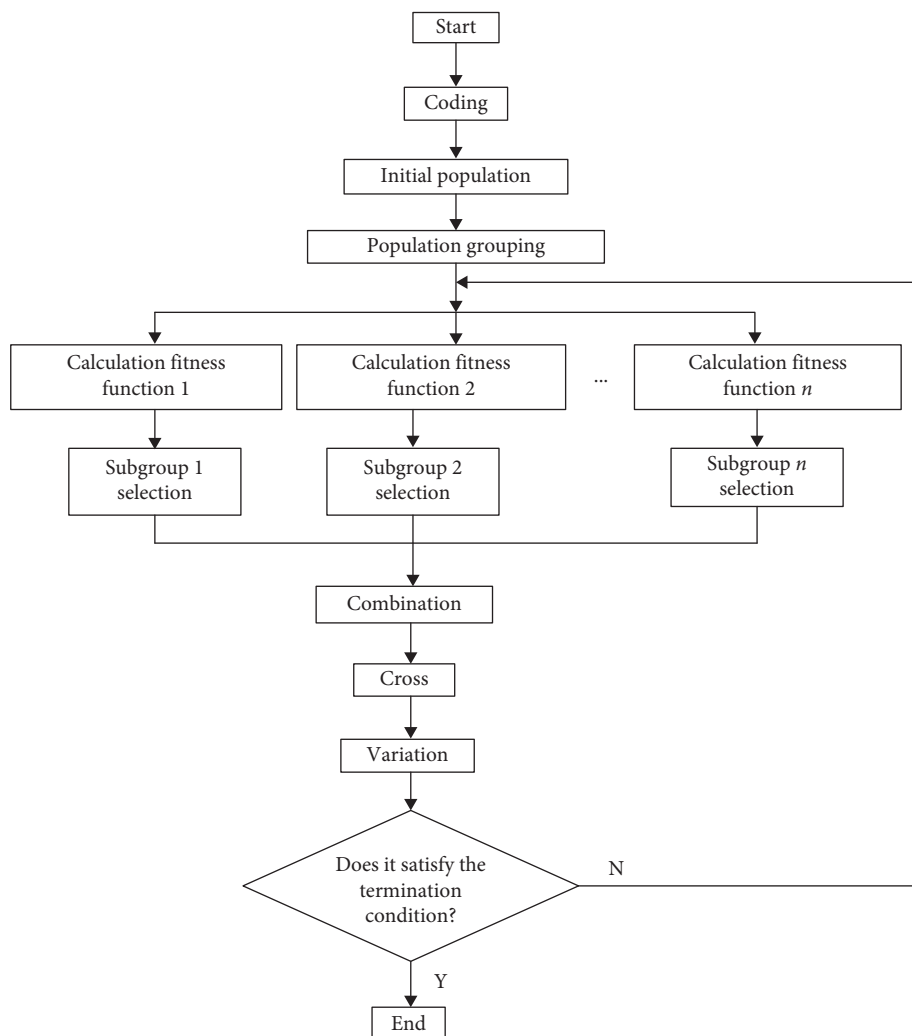


FIGURE 9: Flowchart of parallel selection genetic algorithm after deconstruction and reconstruction.

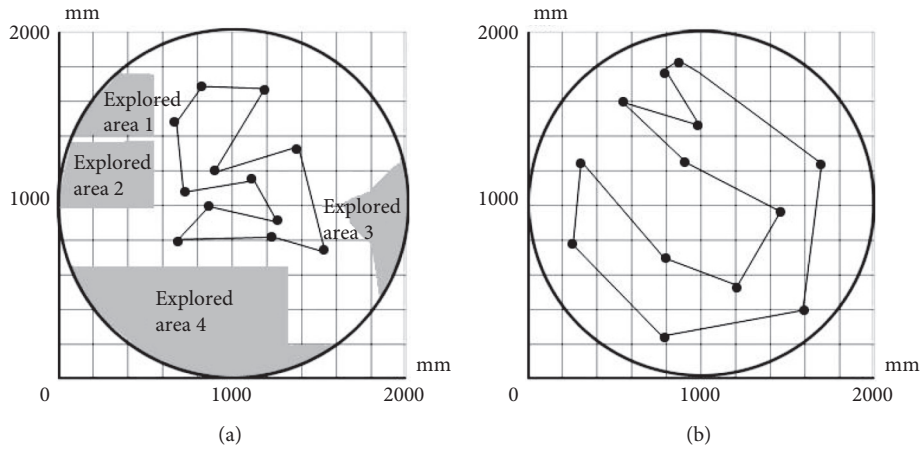


FIGURE 10: Comparison of search area after deconstruction and reconstruction with the random search area. (a) Search trajectory of improved parallel selection genetic algorithm. (b) Search trajectory of random search.

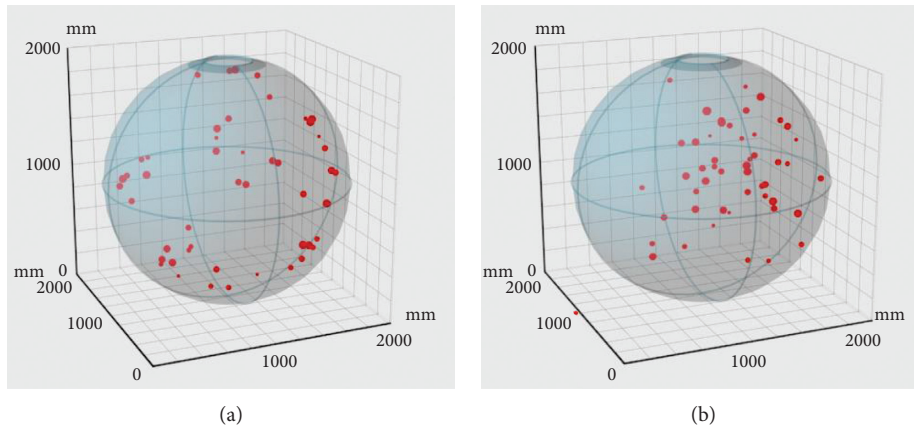


FIGURE 11: Results of multiobjective genetic algorithm for manned submersible cabin layout. (a) Optimization of equipment with low overlapping weight in the cabin. (b) Optimization of equipment with high overlapping weight in the cabin.

equipment and equipment. With the progress of the multiobjective genetic algorithm, the number of collisions (interference amount) in the cabin is gradually reduced, the equipment (small red dots) gradually moves to the center of the sphere and the inner wall, and the distance between equipment and equipment becomes larger, making the activity space of equipment with low overlapping weight scatter on the spherical wall; Figure 11(a) proves that the multiobjective genetic algorithm conforms to the rationality of optimized layout of equipment with low overlapping weight after deconstruction and reconstruction.

As for Figure 11(b), optimization of equipment with high overlapping weight in the cabin, the initial objects are randomly arranged with large activity space, so there is a large amount of interference between equipment and equipment, such as the interference between the activity space of the three submersible pilots and the equipment in the cabin, and between the tasks and the equipment. After deconstruction and reconstruction, with the progress of the multiobjective genetic algorithm, the equipment

(small red dots) gradually moves to the front of the cabin, and the amount of interference decreases after optimization of the distance between equipment and equipment, thus making the equipment with high overlapping weight move in the front of the cabin. Figure 11(b) proves that the multiobjective genetic algorithm conforms to the rationality of the optimized layout of equipment with high overlapping weight after deconstruction and reconstruction.

The results of multiobjective genetic algorithm indicate that when the total interference of equipment in the cabin ( $f_I$ ) is 0, the equipment with low overlapping weight should be fixed on the wall of the cabin; those with high overlapping weight should be fixed in the front of the cabin. In the Jiaolong manned submersible cabin, energy and common tools are arranged in the rear and lower parts of the cabin, the submersible pilots are arranged in the front part of the cabin, and those with high operating and viewing frequency should be placed within the accessibility and visibility range of the 3 submersible pilots.

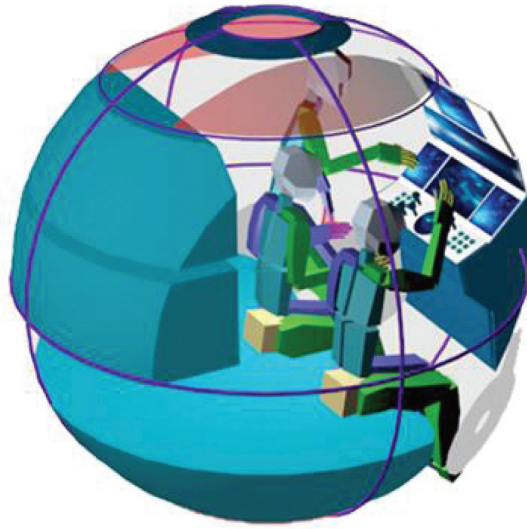


FIGURE 12: Cabin layout optimized by genetic algorithm after deconstruction and reconstruction.

Figure 12 is an effect diagram of cabin layout optimized by genetic algorithm after deconstruction and reconstruction; Pareto optimal solution of the optimized mathematical model is the one obtained with the total interference of equipment  $f_I = 0$ . The human-machine integrated evaluation system for manned submersibles can select a relatively reasonable layout scheme from the optimal solution. As shown in Figure 12, there is no interference in cabin layout optimized by genetic algorithm after deconstruction and reconstruction, indicating a reasonable optimization process.

## 5. Conclusion

The optimization of the cabin layout of manned submersible is of great importance to improve human reliability. In this paper, the manned submersible cabin layout is optimized based on deconstruction and reconstruction thinking from the cross-disciplinary cultural perspective in accordance with the design requirements of manned submersible Jiaolong, combined with ergonomic to constraint conditions on the space layout, so as to establish an optimized mathematical model for cabin layout of manned submersible Jiaolong. After deconstruction and reconstruction, this paper finally obtains a better layout scheme combined with the multiobjective genetic algorithm; the main conclusions are as follows:

- (1) Deconstruction and reconstruction of manned submersible cabin space from an interdisciplinary cultural perspective is a new method of cabin layout optimization, which reflects the advantages of cross-field and multidiscipline combination optimization.
- (2) The manned submersible cabin layout is optimized by combining the aesthetic principles of deconstruction and reconstruction with man-machine ergonomics and other conditions. The same type of

activities and equipment with the same nature are combined to save the cabin space; at the same time, weight combined with multiobjective genetic algorithm is introduced to provide a more scientific scheme for the layout of cabin equipment.

- (3) Based on the aesthetic principle of deconstruction and reconstruction and multiobjective genetic algorithm, the cabin layout optimization model of manned submersible reduces the random search area and finally makes the total interference of cabin equipment approach zero as soon as possible, which can accelerate the progress of the scheme.

However, the optimization mathematical model needs to be further refined in more in-depth subject analysis and calculation tool research. In the next step, static and dynamic dimensions of submarine pilots are added, and virtual simulation and bionic algorithm are applied to improve its engineering practicality.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This study was supported by the National Key Research and Development Program of China “General Design, Construction, and Sea trial of Full Ocean deep Manned Submersible” (2016YFC0300600); Central University Basic Scientific Research Operating Expenses Subsidy Project (31020190504007); and Research on Optimization Design

Method of Deep-Sea Manned Vehicle Cabin under the Support of Composite Simulation Mechanism.

## References

- [1] C. Alppay and N. Bayazit, "An ergonomics based design research method for the arrangement of helicopter flight instrument panels," *Applied Ergonomics*, vol. 51, pp. 85–101, 2015.
- [2] C. Vogt, C. Mergl, and H. Bubb, "Interior layout design of passenger vehicles with RAMSIS," *Human Factors and Ergonomics in Manufacturing*, vol. 15, no. 2, pp. 197–212, 2005.
- [3] Y. Ren, B. Liu, Z. Ding et al., "Research on the current status and development trend of manned submersibles," *Journal of Marine Technology*, vol. 37, no. 2, pp. 114–122, 2018.
- [4] W. Liu, *Application of MDO Method to 7000 m HOV General Design*, Shanghai Jiao Tong University, Shanghai, China, 2007.
- [5] D. Han, H. Han, F. Liu, and X. B. Wang, "Research on 3D layout optimization of HOV cabin," *Ship Engineering*, vol. 35, no. 1, pp. 76–80, 2013.
- [6] D. Chen, Yu Fan, S. Zhang et al., "Manned submersible cabin layout design method under cabin function system constraints and ergonomic constraints," *Chinese Naval Research*, vol. 2, pp. 41–50, 2018.
- [7] Y. Gu, Q. He, X. Fan et al., "Motion planning and intelligent optimization of virtual human arm for narrow space assembly simulation task," *Computer Integrated Manufacturing System*, vol. 22, no. 6, pp. 1447–1455, 2016.
- [8] H. U. Yao, Z. Jiang, Z. Xiong et al., "Warship cabin layout design optimization based on an improved genetic algorithm," *Chinese Naval Research*, vol. 9, no. 1, pp. 20–30, 2014.
- [9] W. Wu, L. Ye, X. Shao, and H. Liu, "Working posture simulation method for virtual human based on multi-objective genetic algorithm," *Computer Integrated Manufacturing System*, vol. 25, no. 1, pp. 155–164, 2018.
- [10] A. Cao, *Multidisciplinary Design Optimization Method and its Application in HOV Design*, Shanghai Jiao Tong University, Shanghai, China, 2008.
- [11] Y. Wang, C. Wang, Z. Ji, and X. G. Zhao, "A study on intelligent layout design of ship cabin," *Shipbuilding in China*, vol. 54, no. 3, pp. 139–146, 2013.
- [12] W. Wang, S. Zhang, and S. Yu, "Image restoration by BP neural based on PSO," *Xibei Gongye Daxue Xuebao/Journal of Northwestern Polytechnical University*, vol. 36, no. 4, pp. 709–714, 2018, in Chinese.
- [13] Y. E. Cong, L. Zong, and D. Chen, "Research and application on virtual design in cabin layout design of deep-sea submersible," *Mechanical Design*, vol. 1, pp. 6–12, 2015.
- [14] J. Jiang, J. Zhao, and X. Zang, "Coordinated multi-robot exploration based on parallelism selection genetic algorithm," *Computer Engineering and Design*, vol. 29, no. 5, pp. 1218–1221, 2008.

## Research Article

# Green Energy Strategic Management for Service of Quality Composition in the Internet of Things Environment

Jianhao Gao 

*School of Accounting Administration, Southwestern University of Finance and Economics, Chengdu, Sichuan 611130, China*

Correspondence should be addressed to Jianhao Gao; 1151202z6007@smail.swufe.edu.cn

Received 21 October 2020; Revised 16 November 2020; Accepted 20 November 2020; Published 30 November 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Jianhao Gao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of Internet of Things (IoT) technology, the energy consumption of service composition in the IoT environment is a key problem to be studied. At present, the problems of service composition in the IoT environment mostly focus on the evaluation research based on quality of service (QoS), ignoring the overall energy consumption in the process of dynamic configuration of service composition. Therefore, we construct the service composition structure for the IoT and propose the QoS evaluation model and energy evaluation model for the service composition in the IoT environment. Considering that the service composition in the Internet of things environment is NP hard, moth algorithm (MFO) is successfully applied to the QoS evaluation model and energy evaluation model. The simulation results reveal that MFO has good optimization effect in the abovementioned models, and the optimization effect of MFO is improved by 8% and 6% compared with the genetic algorithm and particle swarm optimization, so as to realize the green energy strategic management of QoS composition in the environment of IoT.

## 1. Introduction

In recent years, the development of embedded devices has been rapid along with the growth of Internet of Things (IoT) [1–3]. The service-based information system connects the physical reality world and the network virtual world, making the boundary between them gradually blurred [4]. In the IoT, the service-oriented method is widely used and accepted in the development and integration of information [5]. More and more network resources can be freely obtained, such as public data and applications [6]. The traditional information publishing platform is gradually transformed into an open distributed computing infrastructure. With the wide acceptance of service-oriented, service has become the core resource in the network environment [7].

Web services provide the mechanism of service description, service publishing, and service discovery, forming an open, independent, and autonomous distributed environment [8]. In order to ensure the reusability of services, the function of a single service is relatively simple, but in the real application

environment, a simple single service cannot meet the needs of users. When a single service cannot meet the complex needs of users, it is necessary to combine a large number of services with simple functions to form a powerful service that can meet the needs of users. Therefore, scholars hope to create new and more powerful service functions by combining the existing service integration, so as to make full use of the previous resources, expand and extend the original services, fully tap the potential of Web services, and make them play a greater role [9]. From the perspective of task planning, service composition decomposes complex large-scale tasks and then selects atomic services that can complete the subtask for each subtask. With the continuous development of IoT and Web services technology, more and more functions and processes are packaged into standard Web services and published to the Internet, which leads to the exponential growth of the number of services. In order to ensure the interoperability of various information systems, researchers at home and abroad directly apply the service-oriented framework and Web service standards to physical devices. However, it is not appropriate to apply Web service standards directly in the IoT environment [10].



At present, most of the research studies on the quality of service (QoS) composition focus on the optimization of QoS indicators, and some preliminary results have been achieved [11]. However, how to realize the QoS-based green energy strategic management is still an open problem [12]. In reference [13], hybrid integer linear programming (MILP) is proposed for service selection, in which energy consumption is used as an indicator of QoS attributes. However, the time complexity of ILP increases exponentially with the size of the problem. In view of the limitations of integer linear programming (ILP) [14], service selection is formulated to solve the service composition quality problem with multiple constraints. References [15, 16] break through the bottleneck of the classical service composition method which can only select a single candidate service and propose to allow multiple candidate services from a given service class to execute sequentially to improve the QoS of composite services [17]. The advantage of this method is to improve the selection time of the traditional multiconstrained shortest path by introducing the potential path of QoS requirements [18]. In reference [19], service composition is modelled as a path search problem of a directed acyclic graph, and heuristic ant colony algorithm is adopted to search the optimal path in a directed acyclic graph under global constraints [20].

However, the abovementioned research does not fully consider the dynamic nature and the complexity of service state in the optimization process [21, 22]. In reference [23], K-means clustering technology is used to divide candidate services into clusters with roughly the same attributes. These different clusters can be used to determine the best service composition under the condition of satisfying global QoS constraints [24]. Firstly, the QoS value is defined as a discrete random variable with probability quality function, and then, simulated annealing algorithm is used to select service composition satisfying global QoS constraints [25, 26]. The advantage of the abovementioned method is that it can be optimized according to the QoS attributes that users are interested in, for searching the QoS composition path that satisfies the actual needs of users [27, 28]. In references [29, 30], the paper proposes the service composition selection by minimizing the weighted sum of energy consumption and response time.

To solve the abovementioned problems, this paper models the QoS evaluation and energy consumption of the QoS-based composition problem in the IoT environment and applies MFO algorithm to the abovementioned model successfully. Considering the quality of service composition and energy consumption, this paper puts forward the QoS evaluation model and energy evaluation model in the IoT environment and applies MFO (month flame optimization) algorithm to the QoS evaluation model and energy evaluation model successfully. The experimental results show that compared with PSO (particle swarm optimization) and GA (genetic algorithm), MFO can obtain the best QoS and effectively reduce energy consumption, so as to realize the green energy strategic management of service composition quality in the Internet of Things environment. The main contributions of this paper are as follows: (1) most of the

service composition problems in the Internet of Things are focused on the evaluation of quality of service (QoS), ignoring the overall energy consumption in the dynamic configuration of service composition. The service quality evaluation model and energy evaluation model of service composition in the Internet of Things environment are proposed to solve the abovementioned problems; (2) the month algorithm (MFO) is successfully applied to the QoS evaluation model and energy evaluation model. In addition, good optimization effect has been achieved.

The rest of this article is designed as follows. Section 2 gives the structure and model design of QoS composition in IoT. In Section 3, the QoS-oriented green energy management model for IoT is studied. Section 4 conducted simulation experiments and analysed the results. Section 5 summarizes the full text.

## 2. Structure and Model Design of QoS Composition in IoT

How to realize the optimal selection of QoS composition in the IoT environment to meet the actual needs of users for services is an NP hard problem. At present, a large number of literature research mainly focuses on QoS-based service selection. However, with the increase of service categories, each candidate service generates energy consumption, which makes the overall energy consumption increase. Therefore, QoS evaluation and energy consumption evaluation are of great significance for the optimization of service composition. Therefore, structure and model design of QoS composition in IoT is the core research.

*2.1. Multilayer Architecture Design of IoT.* IoT is closely related to a ubiquitous network, and its hierarchical structure is also very similar to a ubiquitous network. The multilayer structure of IoT includes a perception layer, access layer, network layer, support layer, and application layer, which is shown in Figure 1. The perception layer mainly realizes the collection and information processing. The research content of the sensing layer is how to achieve the goal of miniaturization and intellectualization of nodes while saving energy and how to use renewable energy. At the same time, multihop and other research contents of the traditional WSN can be reflected in this layer. The access layer mainly completes the Internet access of all kinds of devices. The network layer is the original Internet, which mainly completes the function of long-distance transmission of information. There are still many contents to be further studied in this layer, such as Internet of Things content distribution technology. The support layer, also known as the middle layer or business layer, mainly completes three parts of functions: the lower needs cognitive network resources and then achieves the purpose of adaptive transmission; its core content is to complete the expression and processing of information and ultimately achieve the purpose of information sharing; this layer also needs to provide a unified interface and virtualization support; virtualization includes computing virtualization and storage virtualization;

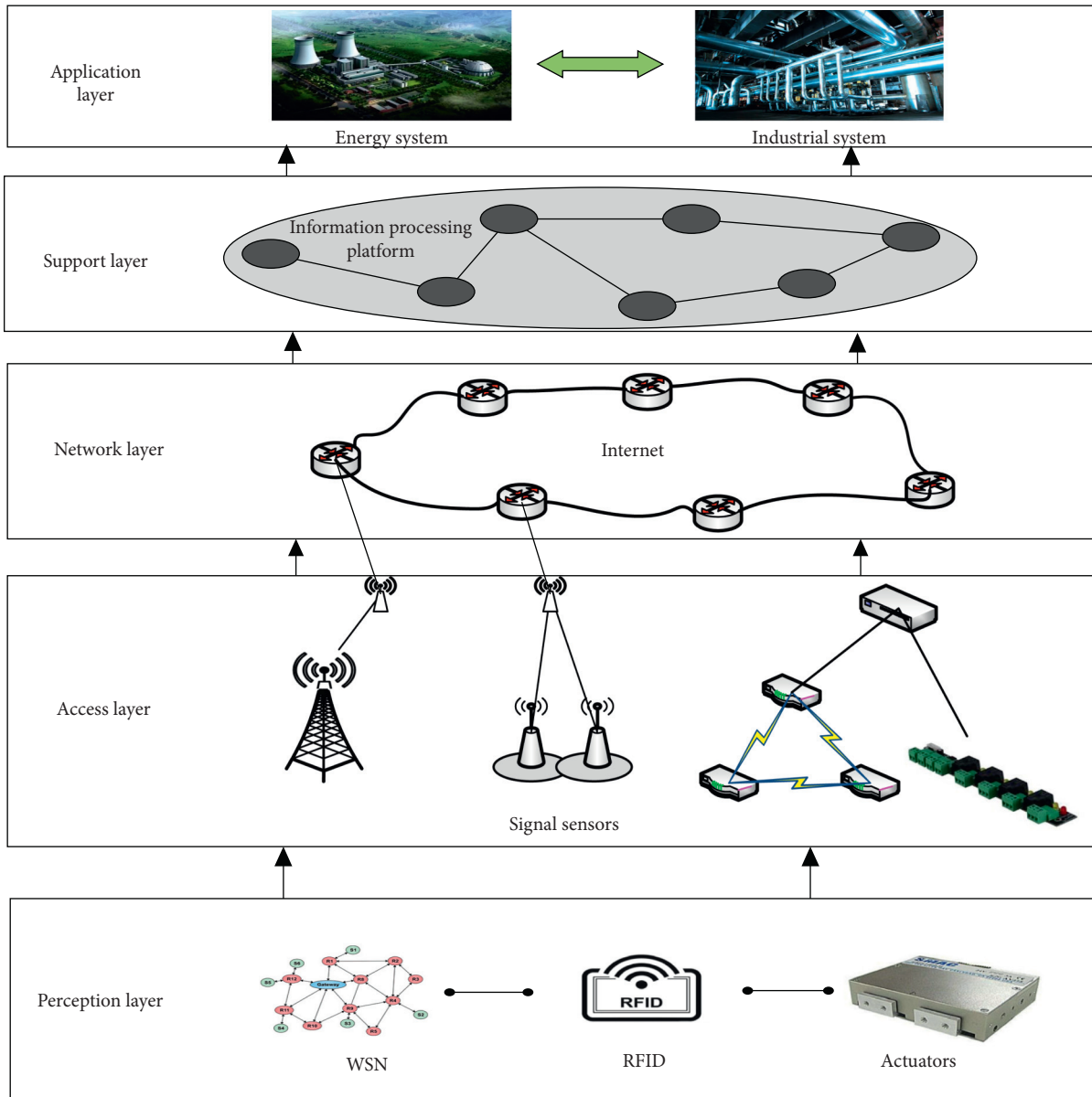


FIGURE 1: Multilayer architecture of IoT.

the more typical technology is Cloud Computing. The application layer is the comprehensive utilization of information to provide help for users.

**2.2. QoS Constraint Model in IoT.** Based on the new features of context adding QoS ontology and semantic IOT services, a QoS constrained service model in IOT is proposed, as shown in Figure 2. Among them, the interaction model shows the relationship between IOT services, service producers, and service consumers; the state model shows the state transition of services; the function model shows the behaviour of services; the three and the relationship between them are marked by the QoS ontology of relevant ontology in the knowledge base.

In the service model based on the Internet of Things shown in Figure 2, the protocol of the underlying sensor may

be quite different from the transmission protocol of the Internet, so there is a home gateway between the two network protocols. Home gateway is responsible for the access of nodes in a small range. Either ID or home protocol are not supported at the same time. In addition, objects may use a variety of physical layer transmission methods, such as wired, WiFi, ZigBee, or Bluetooth, so the home gateway must also have the function of heterogeneous network access. In the case of coexistence of various heterogeneous and same frequency wireless transmission modes, how to reduce the interference between various wireless transmission modes is particularly important, such as the interference problem when ZigBee and WiFi coexist.

After the introduction of object mobility, the network architecture may also need to introduce an access gateway (located between the home gateway and the regional server), which

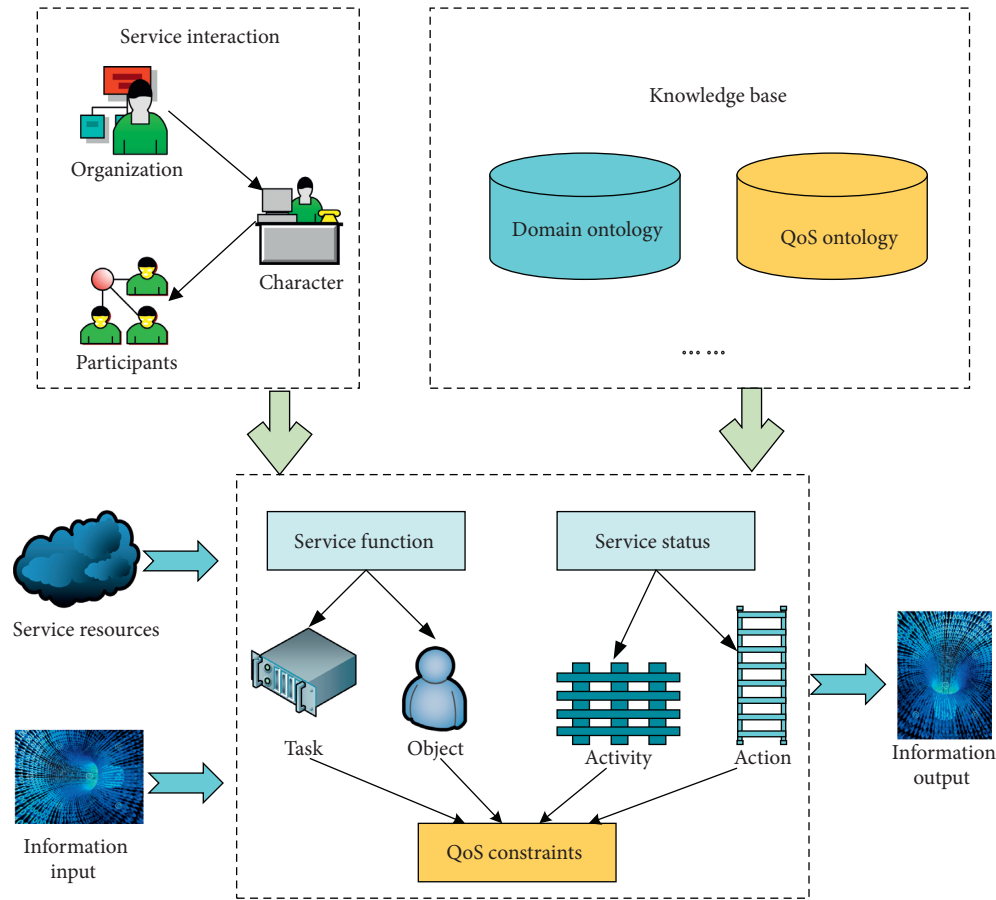


FIGURE 2: Service model based on IoT.

is mainly responsible for the mobility management of nodes in a larger range and the management of existing IOT devices. Of course, if the home gateway is powerful enough, the mobility management of objects can be completed directly in the home gateway under the guidance of a network flattening idea. From the networking technology of the Internet of Things discussed before, the gateway based on DTN is more suitable for architecture of the IoT, so the home gateway discussed here can be implemented in the way of DTN. However, in order to meet some real-time applications (such as actuator control and alarm), the home gateway must have the priority-based data transmission function.

**2.3. Dynamic Composition Framework of Internet of Things Services Based on QoS.** Figure 3 shows the core functions of the dynamic composition framework of QoS-based semantic IoT services; it also shows the dynamic composition model based on QoS and how to optimize the combination mainly including the following modules:

(1) **Service preliminary screening module:** the module analyses the user's direct request for the target service, and the related domain ontology forms the standard service request description file, carries on the preliminary matching with the service in the service library, and obtains the candidate service set.

(2) **Restricted service description file generation module:** the module senses the context information related to users and physical devices through the sensing devices around the physical devices mapped by users and candidate service sets. With the support of context-added QoS ontology, the module calculates the impact of context information changes on the QoS information of services required by users and services provided by physical devices and forms a standard Q without a semantic conflict service description file of OS restriction.

(3) **QoS-based service selection module:** after dimensionless processing of QoS parameter values of the request service description file and each service description file in the candidate service set, atomic matching and aggregate matching are carried out, respectively, and compared with corresponding atomic threshold and aggregation threshold, and one or more subservices that meet some or all requirements of users are selected to form the service set to be combined.

(4) **Service composition module based on graph search:** the module dynamically constructs the directed graph of the service and the input/output data of the service set to be composed and obtains one or more new services that meet the specific conditions. The

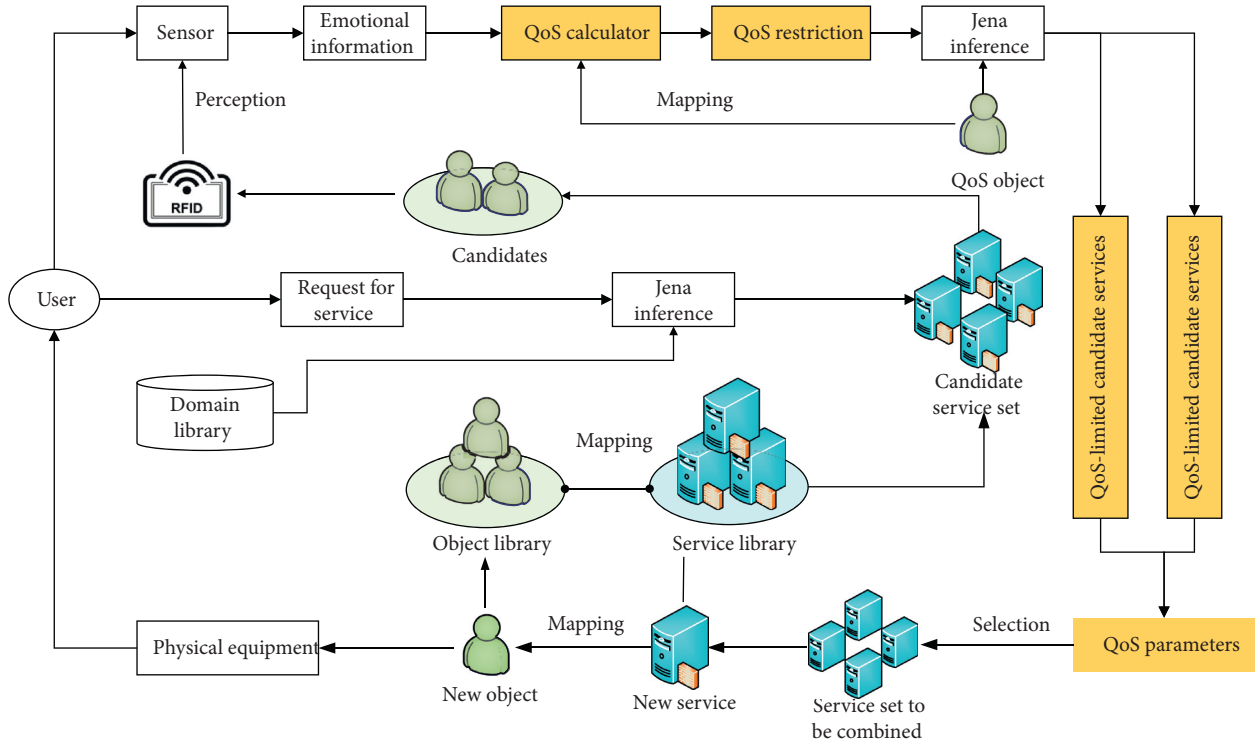


FIGURE 3: Dynamic composition diagram of IoT services based on QoS.

method of combining the depth first algorithm with the breadth first algorithm is used to select the most suitable service.

- (5) Service mapping module based on the object net: through the service ID of each subservice in the new service, the module maps the new service to the corresponding virtual object to form a virtual object network. Finally, the service results are fed back to the user through the corresponding physical settings of the object network.

### 3. QoS-Oriented Green Energy Management Model for Internet of Things

**3.1. QoS Evaluation Model in the Environment of IoT.** Assuming that a composite service can be composed of  $N$  services with different functions, all services can be divided into  $n$  classes according to their corresponding functions, which are called service classes. Assuming that there are  $m$  atomic services in each service class, each service is called a candidate service. The core processing idea of the QoS-driven service composition model is based on the corresponding QoS attribute requirements, and a candidate service is selected from each service class, so as to find a service sequence that can make the QoS attribute optimal or better.

We pay attention to the QoS evaluation model in the Internet of Things environment.  $T = \{T_1, T_2 \dots T_i \dots T_D\}$  is defined as a complex Internet of Things service, where  $D$  means the service categories for a service.

Each service category  $T_i$  contains  $C_i$  specific service  $S_i^j$ , where  $i$  is the service category,  $j$  means the number of following services, and  $j = 1, 2, \dots, C_i$ , the set of QoS metrics used to evaluate each candidate service.

Figure 4 describes the main process of service composition of IoT. As can be seen from Figure 1, the goal of IoT service composition is to select the best path to satisfy the service requirements of customers. Theoretically, there are  $\prod_{j=1}^D C_j$  service composition paths. The path of QoS composition simplifies as  $x = \{x_1, x_2, x_i, \dots, x_D\}$ , where  $x_i$  is the candidate service for executing the  $i$ -th service category, where  $1 \leq i \leq C_i$ . Generally, according to several reduction principles, serial, parallel, and circular structures can be converted into sequential structures when evaluating paths with QoS metrics.

We suppose that the QoS parameter value of each service  $CS_i$  is  $q\{CS_i^j\} = \{q_1, q_2, q_3\}$ , that is,  $q\{CS_i^j\} = \{t, c, r\}$ . The QoS is defined as  $Q = Q\{Q_1, Q_2, Q_3\} = \{T, C, R\}$ , and indicators are estimated, considering the specific rules and the characteristics of the QoS indicator. The target is defined as

$$\min f_1(x_1) = w_1 T(x_1) + w_2 C(x_1) + w_3 R(x_1),$$

$$s.t. \begin{cases} 1 \leq j \leq C_i, \\ 1 \leq i \leq D, \\ \sum_{i=1}^D W_i = 1, \end{cases} \quad (1)$$

where  $\{w_1, w_2, w_3\}$  corresponds to the weight of each QoS indicator, which is set to 1/3 in this paper.

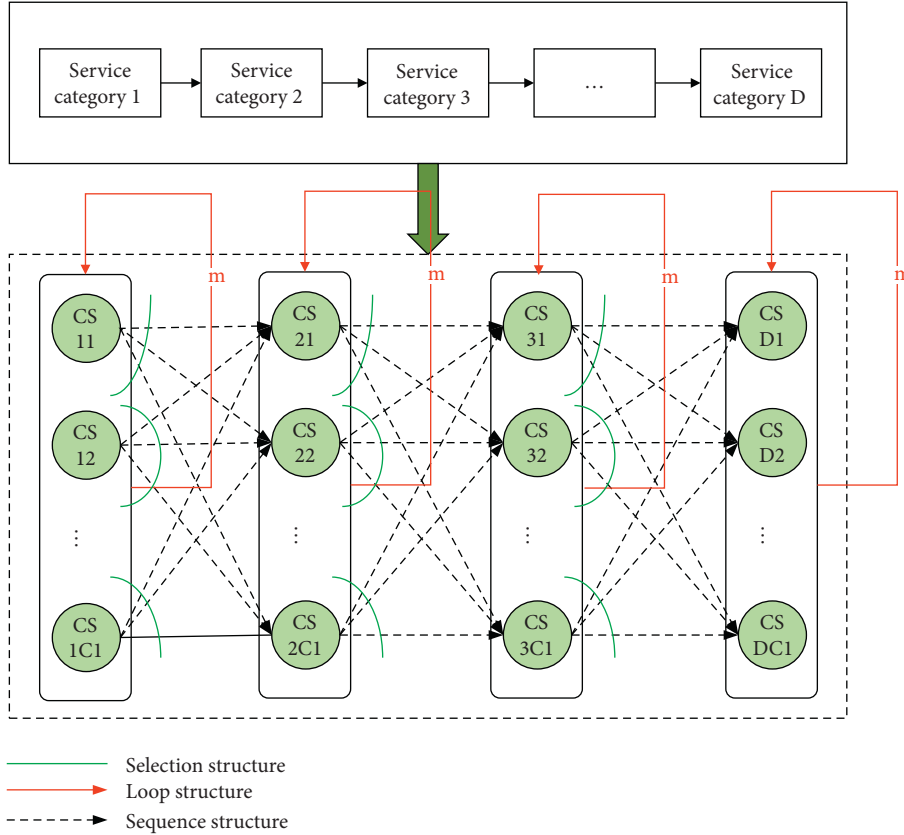


FIGURE 4: The QoS composition process of IoT.

3.2. *Energy Assessment Model Based on IoT.* The objective function of the energy evaluation model is defined as follows:

$$\min f_2(x_2) = \sum_{i=1}^{N_{ES}} \left( \frac{P_i(t)}{\eta_i} C_{run}(i) + P_i(t) C_{om}(i) + P_{Grid}(t) \right), \quad (2)$$

where  $T$  and  $N_{ES}$  are number of system segments and energy services (ES). ES includes various kinds of power generation services, such as photovoltaic (PV) service, wind turbine (WT) service, battery (BAT) service, and fuel cell (fuel cell) service (FC) service, and microturbine (MT) service;  $i$  is the number of ES categories.  $P_i(t)$  is the power generation at time  $t$ , and  $\eta_i$  is the generation efficiency of the  $i$ -th ES.  $C_{run}$  is the operating cost of each ES;  $C_{om}$  is the maintenance cost of each ES.  $C_{Grid}$  is the power price grid, and  $P_{Grid}$  is the exchanged power.

The power balance constraint equation is as follows:

$$\sum_{i=1}^{N_{ES}} P_i(t) + P_{Grid}(t) = P_{Load}(t) + \sum_{n=1}^{N_{ESO}} P_{ESO}(t), \quad (3)$$

where  $P_{Load}(t)$  is the load power at time  $t$  and  $P_{ESO}(t)$  is the discharge power of ES at time  $t$ . Equation (3) illustrates that the IoT-based energy management must meet the total power balance conditions in the whole period.

The power constraint equation is as follows:

$$\begin{cases} P_{ES,min} \leq P_{ES}(t) \leq P_{ES,max}, \\ \Delta P_{ES,down} \leq P_{ES}(t) - P_{ES} \leq \Delta P_{ES,up}, \end{cases} \quad (4)$$

where  $P_{ES,min}$  and  $P_{ES,max}$  are minimum and maximum limits of the ES output;  $\Delta P_{ES,down}$  and  $\Delta P_{ES,up}$  are the maximum reduction and increase of power per unit time of ES.

3.3. *Green Energy Model of Internet of Things by Adopting Moth Algorithm.* Because the QoS composition in IoT is an NP hard problem, the traditional mathematical methods have some problems such as high computational complexity and low efficiency. The intelligent optimization algorithm has better optimization effect in solving the abovementioned problems. The references show that MFO has better optimization effect and higher stability than the traditional intelligent optimization algorithm in single peak function, multimodal reference function, and composite reference function. Therefore, this paper selects MFO to simulate QoS evaluation model and energy evaluation model.

The population size of moth is  $n$ , and its flight space is  $d$  dimension, which is the same as the number of service categories. Initial population  $\{x_1, x_2, \dots, x_n\}$  where  $X_i$  is the position of the moth.  $x_i$  is produced by the following equation:

$$x_i^d = [lb^d + (ub^d - lb^d) \times \text{rand}(0, 1)], \quad (5)$$

where  $ub$  and  $lb^d$  are the upper and lower limits of  $d$ -dimension, respectively, and  $\text{rand}(0, 1)$  is a value randomly generated between  $(0, 1)$ . FMO defines the moth population and the corresponding fitness value, the flame population, and the corresponding fitness value by combining the specific trajectory of moth tending to artificial light. By updating the moth population in the form of helix and calculating the fitness value of the evolutionary moth population, the optimal fitness value and corresponding solution are obtained. In this algorithm, the update of moth position is as shown in the following equations:

$$s(M_i, F_j) = D_i \times e^{bt} \times \cos(2\pi t) + F_j, \quad (6)$$

$$D_i = |F_j - M_i|, \quad (7)$$

where  $F_j$  means the moth position,  $M_i$  is the flame position, and  $D_i$  shows the distance between the position of the moth and the position of the flame. It can be seen from formula (7) that the position update of the moth is relative to the position of the flame. Let  $t$  in formula (8) be a random digit between  $[r, 1]$  in the iterative process. In this way, the moth can make better use of the corresponding flame in the iteration process.

## 4. Simulation Results and Analysis

**4.1. Simulation Parameter Setting.** Due to the lack of a common data set, the service composition problem in the IoT environment is simulated by using the data set with composite QoS attributes according to the references. In order to verify the effectiveness of the algorithm, MFO is compared with PSO and GA. Software environment: Microsoft Windows 10 professional edition (64 bit), version number 16299, Matlab r2015a. Hardware environment: inter (R) core (TM) i7-6700 CPU @ 3.40 GHz, 3408 mhz, 4 cores, 8 logical processing units, 8.00 gb physical memory, 1 t hard disk.

The experimental parameters are set as follows: for the QoS evaluation model, the QoS attribute value of each service is randomly generated, and the service category  $D$  and candidate service  $C$  increase from a small value. For the energy consumption evaluation model, the duration of sample interval  $\Delta t$  is 1 h, and total duration was 24 hours.  $i$  is the category of energy services, and the initial population  $N$  is the number of candidate services. In the experiment, two cases were selected:  $i$  fixed at 72,  $N$  increased from 100 to 1000;  $n$  fixed at 100,  $I$  changed from 72 to 720. The parameters of GA are shown as follows: the population number is 100, iterations are 200, fitness normalized elimination acceleration index is 2, crossover probability is 0.8, and mutation probability is 0.05. The initial position is random. The main parameters of PSO are as follows: the max iterations are 800; number of independent variables of objective function is 2; particle swarm size is 50; and individual learning factor and social factor of each particle are 2.

**4.2. Validation of the QoS Evaluation Model.** In order to verify the change of an individual's maximum fitness with population algebra, firstly, the number of service classes in

composite service is set to 45, and the number of candidate services in each service class is 35. The corresponding test data are generated by the range of QoS attribute generation.

As is revealed from Figure 5, the performance of MFO algorithm is obviously better than that of GA and PSO algorithms under different number of alternative services. The difference is smaller and more stable. We can obtain that MFO has better global optimization ability and can find the global optimal solution of the service composition problem with greater probability.

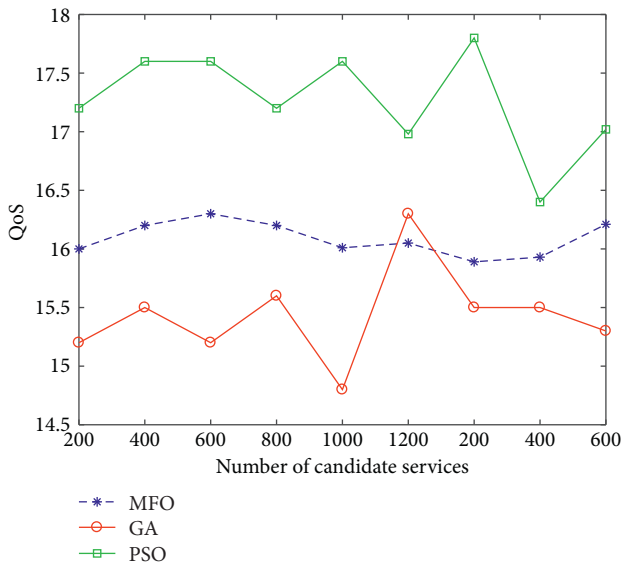
At the same time, the fitness function value of the QoS evaluation model in Figure 5(b) decreases with the increase in the number of iterations. With the increase in iteration times, the error of fitness function gradually decreases, which shows that the optimization efficiency of MFO is higher than that of particle swarm optimization and genetic algorithm. MFO algorithm can reach the convergence state after 1000 iterations, which is more efficient than 1300 and 2200 times of PSO and GA. This also indirectly shows that the optimization strategy based on moth algorithm is better than particle swarm optimization and genetic algorithm. The sharp increase of GA results is mainly caused by three reasons: the multisolution genetic algorithm does not converge, the initial feasible solution has problems, and the evaluation function is not distinguished enough.

**4.3. Validation of the Energy Evaluation Model.** In order to obtain the best combination of QoS, the method of choosing proper parameters is the same as that in the QoS model. In this simulation, MFO, PSO, and GA are selected to estimate the energy consumption under different parameters. Figure 6(a) shows that the optimization effect of MFO on the energy consumption evaluation model is significantly better than PSO and GA.

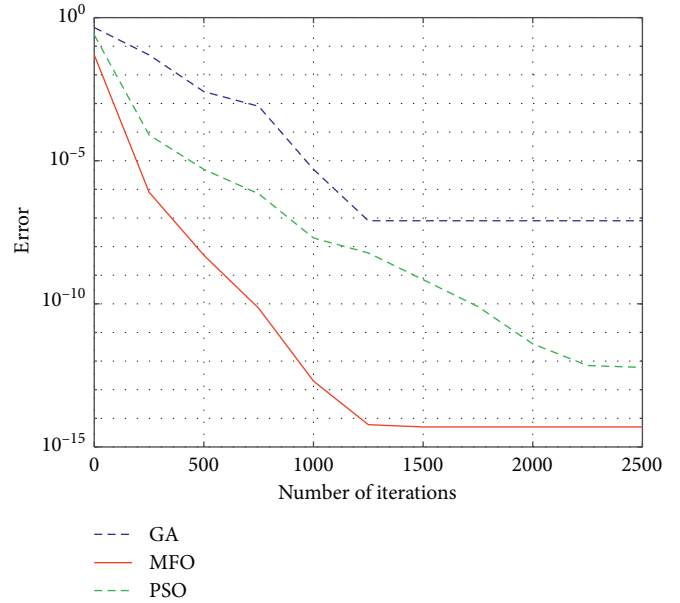
In Figure 6(b), simulation results reveal that the energy consumption of MFO-based service combination in the Internet of Things environment is significantly less than that of PSO and GA algorithm. In addition, Figure 6(b) shows that, with the increase of candidate services, energy consumption also increases because each service category selection will generate energy consumption, and a linear relationship existed. Also, the optimization effect of MFO is improved by 8% and 6% compared with the genetic algorithm and particle swarm optimization.

**4.4. Comparison of QoS Energy Consumption in Internet of Things.** In order to verify the design advantages, we compare the performance and energy utilization of the hierarchical management data centre and the UN optimized hybrid data centre. We use professional precision power and energy meters to detect the energy consumption and power consumption of the server in real time and monitor the number of requests waiting for the queue to accumulate to analyse its performance.

Figure 7 shows the change in tasks number in queue with time. The larger the value, the more serious the task accumulation, the lower the performance of the data centre, and the more difficult it is to meet the computing requirements.

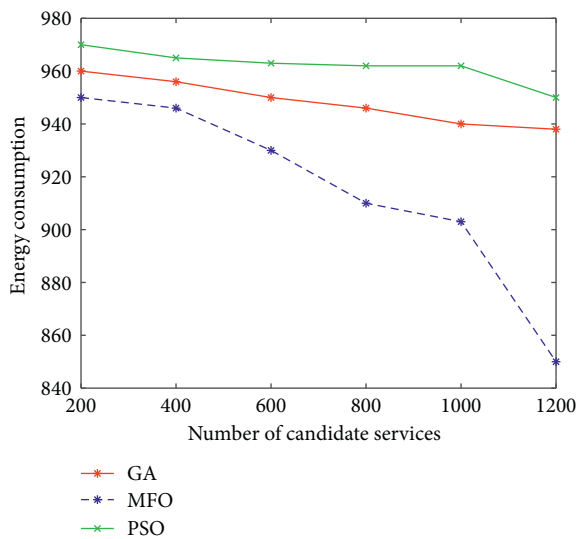


(a)

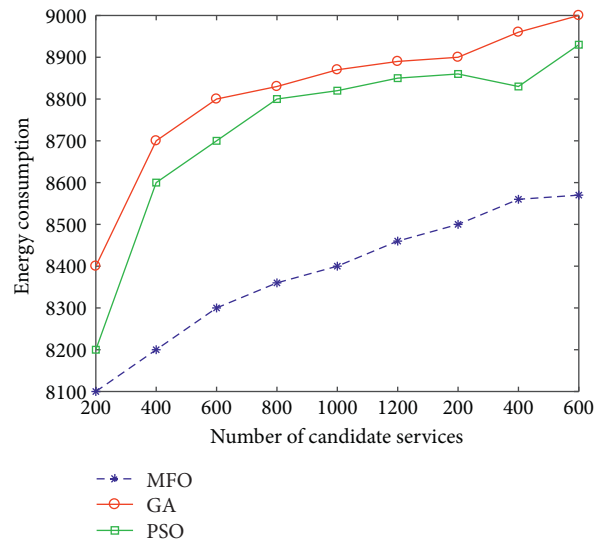


(b)

FIGURE 5: Relationship between QoS and number of services. (a) QoS changes with the number of candidates; (b) error change with the number of iterations.



(a)



(b)

FIGURE 6: Relationship between energy consumption and services quantity. (a) Energy consumption under services quantity. (b) Energy consumption under different service categories.

Figures 7(a) and 7(b) shows the change in the length of the task waiting queue with time under the traditional and QoS-oriented strategies, respectively. The traditional task-scheduling mode lacks the control and allocation of random arriving tasks, so the data centre has the phenomenon of uneven task allocation and certain task accumulation. Compared with Figure 7(a), the tasks in the waiting queue in Figure 7(b) can be very fast. On the one hand, it shows that the hierarchical organizational structure can improve the performance of the hybrid data centre, that is, the waiting time of tasks; on the other hand, it also shows that there is free space for

computing resources in the case of layering. Even if some computing resources are shut down, some of them may still meet the computing needs, and there is a potential to trade performance for energy efficiency.

In Figure 8, we compare the overall energy-saving effect of the data centre under different sleep states. We directly measured that the power consumption of all servers (hybrid data centre system) is about 8% of the maximum power in Figure 8, the ordinate represents the percentage of energy saved under  $4 \times 3$  different dynamic sleep strategies under QoS positioning, and the abscissa

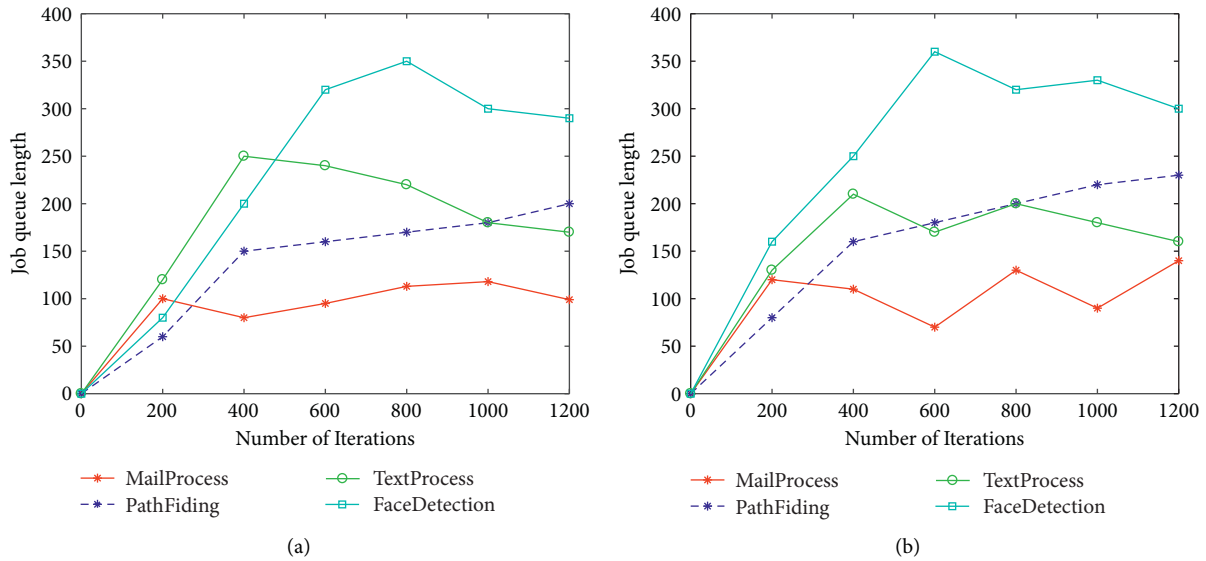


FIGURE 7: Performance monitoring results. (a) QoS-oriented energy consumption. (b) Traditional energy consumption.

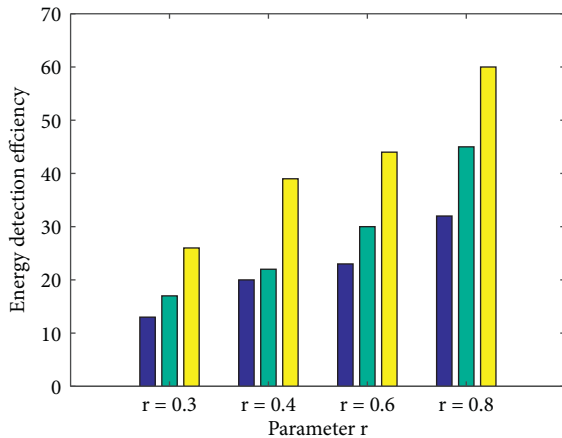


FIGURE 8: Energy consumption monitoring results under different parameters.

represents the sleep ratio of servers in the central data centre. It can be seen that, with the growth of dynamic sleep time, the data centre can save more energy, but this energy efficiency advantage is also affected by the data centre performance constrained, that is, when the closed computing resources are too much or the data centre computing resources are not enough, leading to a serious decline in the QoS. In the experiment, moderate load integration and collective sleep strategy for all levels of IOT data centre can bring more than 12% energy saving. In conclusion, in view of the IOT environment, the hierarchical organization structure of service composition can be realized reasonable management and planning of computing resources, and the reasonable collective sleep strategy in this hierarchical structure can further improve energy efficiency with low performance impact.

## 5. Conclusions

In this paper, aiming at the deficiency of traditional methods only based on QoS evaluation to optimize services in the IoT

environment, this paper proposes a QoS evaluation model and energy evaluation model of QoS-based composition and successfully applies MFO to the QoS evaluation model and energy evaluation model. Simulation results reveal that MFO algorithm has more remarkable optimization effect than PSO and GA. Therefore, when MFO is used to solve the service composition problem, it can not only satisfy the requirements for QoS in the Internet of Things but also achieve optimal energy efficiency. It is proved that MFO has better large-scale processing ability and global optimization ability than traditional genetic algorithm and PSO and can find the global optimal solution of the QoS composition problem with greater probability. In the next step of research, we can obtain real data sets from experiments to evaluate the QoS of composition and energy consumption under the IoT. In addition, MFO has strong performance, but it is possible to fall into local optimum. Therefore, we can consider combining MFO with other intelligent optimization algorithms to improve MFO for overcoming the local optimization for the QoS problem.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no known conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

## References

- [1] W. Ejaz, M. Naeem, A. Shahid, M. Jo, and A. Anpalagan, "Efficient energy management for the internet of things in smart cities," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 84–91, 2017.
- [2] I. Joe and M. Shin, "Energy management algorithm for solar-powered energy harvesting wireless sensor node for internet



- of things,” *Iet Communications*, vol. 10, no. 12, pp. 1508–1521, 2016.
- [3] M. Chinmaya, M. Akshaya, and L. Victor, “Energy management in smart cities based on internet of things: peak demand reduction and energy savings,” *Sensors*, vol. 17, no. 12, p. 2812, 2017.
  - [4] M. Vangelis and D. Haris, “An advanced IoT-based system for intelligent energy management in buildings,” *Sensors*, vol. 18, no. 2, p. 610, 2018.
  - [5] W. T. Cho, Y. W. Ma, and Y. M. Huang, “A smart socket-based multiple home appliance recognition approach over IoT architecture,” *Journal of Internet Technology*, vol. 16, no. 7, pp. 1227–1238, 2015.
  - [6] B. Muhammad, G. Jakub, O. Andrzej et al., “Energy flexometer: transactive energy-based internet of things technology,” *Energies*, vol. 11, no. 3, p. 568, 2018.
  - [7] H.-W. Kim, J. H. Park, and Y.-S. Jeong, “Efficient resource management scheme for storage processing in cloud infrastructure with internet of things,” *Wireless Personal Communications*, vol. 91, no. 4, pp. 1635–1651, 2016.
  - [8] W. J. Shyr, L. W. Zeng, C. K. Lin et al., “Application of an energy management system via the internet of things on a university campus,” *Journal of Periodontology*, vol. 38, no. 3, pp. 1527–1534, 2017.
  - [9] H. Elhammouti, E. Sabir, M. Benjillali, H. Tembine, and L. Echabbi, “Self-organized connected objects: rethinking QoS provisioning for IoT services,” *IEEE Communications Magazine*, vol. 55, no. 9, pp. 41–47, 2017.
  - [10] N. Kumar, S. Zeadally, and S. C. S. Misra, “Mobile cloud networking for efficient energy management in smart grid cyber-physical systems,” *IEEE Wireless Communications*, vol. 23, no. 5, pp. 100–108, 2016.
  - [11] Y. Liu, C. Yang, L. Jing, Y. Zhang, and S. Xie, “Intelligent edge computing for IoT-based energy management in smart cities,” *IEEE Network*, vol. 33, no. 2, pp. 111–117, 2019.
  - [12] N. Abuzainab, W. Saad, C. S. Hong, and H. V. Poor, “Cognitive hierarchy theory for distributed resource allocation in the internet of things,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 12, pp. 7687–7702, 2017.
  - [13] E. Curry, S. Hasan, C. Kouroupetroglou, U. Ul Hassan, W. Fabritius, and W. Derguech, “Internet of things enhanced user experience for smart water and energy management,” *IEEE Internet Computing*, vol. 22, no. 1, pp. 18–28, 2018.
  - [14] S. Derguech, S.-Y. Joo, and S. Silvestri, “Managing contingencies in smart grids via the internet of things,” *IEEE Transactions on Smart Grid*, vol. 7, no. 4, pp. 2134–2141, 2016.
  - [15] O. Bates and A. Friday, “Beyond data in the smart city: repurposing existing campus IoT,” *IEEE Pervasive Computing*, vol. 16, no. 2, pp. 54–60, 2017.
  - [16] W. J. Shyr, C. F. Chiou, F. C. Yang, and P.-C. Li, “Energy management competency development based on the internet of things (IOT),” *International Journal of Engineering Education*, vol. 33, no. 4, pp. 1380–1385, 2017.
  - [17] F. Eleni, Z. Anastasios, T.-S. Fernando et al., “Providing personalized energy management and awareness services for energy efficiency in smart buildings,” *Sensors*, vol. 17, no. 9, p. 2054, 2017.
  - [18] S. Forsstrom and T. Kanter, “Continuously changing information on a global scale and its impact for the internet-of-things,” *Mobile Networks & Applications*, vol. 19, no. 1, pp. 33–44, 2014.
  - [19] M. R. Abdmeziem, D. Tandjaoui, and I. Romdhani, “Light-weighted and energy-aware MIKEY-Ticket for e-health applications in the context of internet of things,” *International Journal of Sensor Networks*, vol. 26, no. 4, p. 227, 2017.
  - [20] J. Zhang, X. Qu, and A. K. Sangaiyah, “A study of green development mode and total factor productivity of the food industry based on the industrial internet of things,” *IEEE Communications Magazine*, vol. 56, no. 5, pp. 72–78, 2018.
  - [21] F. Tao, Y. Zuo, L. D. Xu, L. Lv, and L. Zhang, “Internet of things and BOM-based life cycle assessment of energy-saving and emission-reduction of products,” *IEEE Transactions on Industrial Informatics*, vol. 10, no. 2, pp. 1252–1261, 2014.
  - [22] H. Yang, A. Alphones, W.-D. Zhong, C. Chen, and X. Xie, “Learning-based energy-efficient resource management by heterogeneous RF/VLC for ultra-reliable low-latency industrial IoT networks,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 8, pp. 5565–5576, 2020.
  - [23] M. V. Moreno, M. A. Zamora, and A. F. Skarmeta, “User-centric smart buildings for energy sustainable smart cities,” *Transactions on Emerging Telecommunications Technologies*, vol. 25, no. 1, pp. 41–55, 2014.
  - [24] E. Oh and S.-Y. Son, “A framework for consumer electronics as a service (CEaaS): a case of clustered energy storage systems,” *IEEE Transactions on Consumer Electronics*, vol. 63, no. 2, pp. 162–168, 2017.
  - [25] L. Ren, L. Zhang, F. Tao, C. Zhao, X. Chai, and X. Zhao, “Cloud manufacturing: from concept to practice,” *Enterprise Information Systems*, vol. 9, no. 1–2, pp. 186–209, 2015.
  - [26] Z. Wu, “An empirical study of the accessibility of web references in two Chinese academic journals,” *Scientometrics*, vol. 78, no. 3, pp. 481–503, 2009.
  - [27] S. Mathaba, M. Adigun, J. Oladosu, and O. Oki, “On the use of the internet of things and Web 2.0 in inventory management,” *Journal of Intelligent & Fuzzy Systems*, vol. 32, no. 4, pp. 3091–3101, 2017.
  - [28] A. Alnoman and A. Anpalagan, “Towards the fulfillment of 5G network requirements: technologies and challenges,” *Telecommunication Systems*, vol. 65, no. 1, pp. 1–16, 2017.
  - [29] S. K. Roy, S. Misra, and N. S. Raghuvanshi, “SensPnP: seamless integration of heterogeneous sensors with IoT devices,” *IEEE Transactions on Consumer Electronics*, vol. 65, no. 2, pp. 205–214, 2019.
  - [30] X.-C. Hao, M.-J. Xin, and X.-Y. Ru, “EAPOR: a distributed, energy-aware topology control algorithm based path-obstacle-remove model for WSN,” *Wireless Personal Communications*, vol. 80, no. 2, pp. 671–692, 2015.

## Research Article

# Visualization of Information Retrieval in Smart Library Based on Virtual Reality Technology

**Shulin Fang** 

*Xi'an Academy of Fine Arts, Xi'an, Shaanxi 710065, China*

Correspondence should be addressed to Shulin Fang; 50041@xafa.edu.cn

Received 23 October 2020; Revised 16 November 2020; Accepted 18 November 2020; Published 29 November 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Shulin Fang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Starting from the virtual reality technology, the characteristics of its most suitable combination with the library are explored, so as to lay a foundation in theory and practice to promote the development of virtual reality in the library. In the concentration camp of the latest advanced technology, the relevant technologies used in the various levels of models in the smart library are extracted, and their functional principles and applications are systematically introduced; Chapter 4 builds the level of the smart library information retrieval technology model. The structure diagram, various levels of functions, and related smart service models are discussed. Research is on context construction of extended resources of knowledge service in smart libraries. The elements and content of the resource context are introduced, and the strategy of constructing the resource context for the extension of the knowledge service of the smart library is proposed. Research is on the construction of the context of the extension and interconnection of knowledge services in smart libraries. The relevant elements and contents of the technical context and the spatial context are introduced, and a strategy for constructing a connected context for the extension of the knowledge service of a smart library is proposed. Research is on the context construction of knowledge service extension of smart library. The elements and contents of the service context are introduced, and the strategy of constructing the service context of the knowledge service extension of the smart library is proposed. A visual model of information retrieval is constructed. The model integrates the core steps of the information visualization process and introduces the information space and functional system (navigation, organization, indexing, and retrieval) in the information construction theory into it and through five mapping layers (functional space mapping, visualization space mapping, visualization mapping, view mapping, and interactive mapping) and six spatial layers (role space, information space, functional space, visualized information space, visualized object space, and visualized view space), which describe visualization applications that target the user experience model in information construction build process.

## 1. Introduction

With the advent of the big data era, a large amount of data continues to emerge, and there is a phenomenon of abundant information but there is a lack of useful information. Information overload has caused problems such as lack of information and a significant reduction in user information utilization. As an indispensable part of society, libraries should bring forth the new and avoid working behind closed doors. They need to adapt to the development of the times and integrate the latest technology into the library to achieve personalized services for users. But no matter how the library changes and how the technology is updated alternately, its core value of “people-oriented” has

not changed. How can the library realize the transformation from the passive and stereotyped model to the intelligent service model in the new intelligent environment and realize the transformation from single-library service to cross-library service? The construction of the technical model of China's smart library studied in this paper provides technical support for the above problems. Relying on the construction of this model, the library can realize transparent management of user services, mine users' useful information needs, and intelligently predict user personalized services trend, accurately integrating the value density of library management, scientific services, and librarian decision-making.

Research on smart libraries can suggest that it is a difficult and tedious task for readers to accurately find the

books and materials they need in the library environment, but location-aware technology can help readers solve such problems and achieve the purpose of accurate search [1, 2]; it is proposed that a smart library is a mobile library that can transcend space limitations and can be noticed by people [3, 4]. The fact that a smart library uses a large number of software quality projects to reduce users and libraries is emphasized. The goal of error is in the process of use, classification, configuration, installation, or processing by technical personnel [5]. Research is on the application of data mining technology in smart libraries. Through this interactive platform system, the library uses EAI technology to construct the underlying architecture, exchange heterogeneous systems, data, and platforms, realize seamless integration within the system and between multiple systems, and then uses scene graph, data mining, and data analysis technology which intelligently perceives, mines, captures, analyzes, and integrates information to achieve collaboration and sharing of data processing [6, 7]. In the era of big data, data mining technology is of great significance to the development of libraries. It not only can extract hidden and potentially valuable knowledge information from a large amount of disordered and vague practical application data but also can be used to support a wide range of business smart applications, such as directional marketing [8, 9]. Smart library mobile visual search service architecture diagram, and display mobile visual search services, visualization services, knowledge services, social services, social recommendations, one-stop navigation services, and other service functions in the smart service module [10, 11]. Innovative service models have distinctive features, and these accompanying features have evolved based on new technologies and intelligent facilities. Research is on location-based information push services and private customization and other service models, and the research and innovative development of smart libraries by expanding social networks and expanding publicity camps have increased [12]. Guided by relevance learning theory, service models of context-aware services, social services, cloud services, and mobile information services are constructed [13]. The deep integration of "Internet +" and the library has realized the interconnection between the physical world and the virtual world, making the library rich and diverse, and the library community is also actively studying this direction [14]. Under the background of the deep integration of traditional industries and the Internet, it is proposed that libraries should establish user thinking and Internet thinking to build a smart library. Internet thinking is the technical basis of user thinking, and user thinking is the manifestation of Internet thinking, include two complement segmentation [15, 16]. In the "Internet +" environment, the service content of the smart library is introduced in detail [17]. Information technology plays an important role in the transformation and upgrading of the library. The use of Internet of Things, cloud computing, wearable technology, virtual reality technology, artificial intelligence, and other technologies can realize the optimization of the library's literature, equipment, personnel, and buildings [18, 19]. Based on the problems in the application of the Internet of Things

identification technology in the smart library, the improvement measures for the construction of the smart library are proposed [20, 21]. The three relationships between wearable technology and smart library, the specific application of wearable technology in reader navigation, and help for disadvantaged groups and personalized services are analyzed [22, 23]. Using the effect of virtual reality to present pictures, models, or videos in the book, the reading method and interest are increased [24, 25]. A three-dimensional book of virtual reality is made, and the pictures in the book are carried out by using mobile devices. The scanning function can present the actual scene of the corresponding 3D model, change the interactive way of parent-child reading, and promote the establishment of the reading sharing relationship between parent and child [26]. Mobile learning is most effective only when it connects the real environment with related resources [27, 28]. Mobile AR technology provides a channel for the establishment of this connection, and library resources make the connection possible. In the research of the library personalized service system based on virtual reality, there is a detailed introduction to the realization of personalized service related functions, including the following: real-time scanning of QR codes, real-time calculation of projection matrix, projection of three-dimensional objects and location-based book retrieval and book recommendation functions of [29, 30]. In the research on the application of virtual reality technology in the library personalized service platform, it is found that the library information browsing system based on QR code and virtual reality can present different information interfaces to different users according to the information displayed by scanning the QR code. It guides readers to browse the information of books on the bookshelf in an intuitive way and at the same time recommends various resources of interest to readers [31]. The Android-based virtual reality library service platform is a platform that combines new technologies such as the mobile Internet and data mining technology to provide users with personalized services such as book query, book recommendation, and book location navigation [32]. The impact of various frontier technology developments is on libraries, including virtual reality technology. The ideas and countermeasures to innovate the library service model for reference from the Internet are put forward [33]. In the exploration of the application of virtual reality technology in the library, starting from the practical application direction of the library, the foundation of virtual reality is introduced in detail, the important role of this technology in the library business is explored, and the effect of virtual reality in the library is initially explored [34]. The application status of mobile virtual reality technology in the library, combined with the characteristics of various services of the library, is analyzed and the application prospects of mobile virtual reality technology in the library are also explored in order to further utilize the mobile virtual reality technology in the library. The application value in the library provides reference [35]. From the perspective of bookshelf and resource integration, virtual reality multimedia books, virtual reality library navigation, virtual reality optical character recognition, and virtual reality personalized

services, the application of mobile virtual reality technology in modern libraries is discussed [36, 37]. Through research and analysis, the necessity of using virtual reality technology in smart libraries is analyzed, and the application form, workflow, and image recognition mechanism of virtual reality in smart libraries are pointed out, and the innovative service mode of smart libraries based on virtual reality technology is explored [38]. The current situation of library business at this stage is analyzed, and, according to the characteristics of virtual reality, the application of virtual reality technology in intelligent libraries from three aspects is discussed: library personalized intelligent navigation service, resource integration service, and personalized recommendation push service [39, 40].

In view of the research on the information retrieval technology model of the smart library, a detailed analysis is made on the specific construction content of the smart perception layer, network transmission layer, data resource layer, and smart application layer. This article mainly focuses on personalized and scenario-based recommendation, and virtual research is carried out on six service dimensions of reality, multimedia, smart space, and visualization. From a new perspective, virtual reality technology has been widely popularized in certain fields and has brought about certain benefits to those fields. One of the evaluation criteria for the effect of the integration of new technology is whether it has improved the industry, and the application of virtual reality in the library is consolidated through various practical improvement effects, so as to lay the foundation for the popularization of the library industry, as well as theory and practice to be prepared for in-depth research. It also proposes a strategy for constructing a resource context for the extension of knowledge services in a smart library to introduce the elements and content of a resource context, a strategy for constructing a context for the extension of knowledge services in a smart library, and a strategy for constructing a context for extension services of a smart library. The application of augmented reality, virtual reality, and integrated reality provides a new way for libraries to carry out innovative services and a new perspective for libraries to better serve readers. With the passage of time, the technologies of virtual reality, augmented reality, and integrated reality will continue to develop, and the applications full of surprises will continue to change the way readers work, communicate, and entertain themselves and further expand the functions of libraries, so as to promote the construction and development of smart libraries.

## 2. Information Retrieval Model of Smart Library Based on Virtual Reality

With the current rapid development of intelligence and Internet of Things technology, a technical platform has been created for the bold use of virtual reality technology in libraries, and it has also provided technical innovation and service concept transformation for the diversified transformation of library services. Virtual reality technology is to reflect virtual information into the real world with the help of computer processing technology to realize the integration

of virtual objects, scenes, actions, and other objects and apply them to real scenes. The introduction of virtual reality technology in the smart library creates a comfortable smart virtual space for readers, allowing readers to enhance their desire to enjoy library services in the context of services. Virtual reality technology is a special form of reality technology. It has the characteristics of strong interaction, integration of virtuality and reality, and three-dimensional positioning. It introduces three-dimensional registration and virtual compatibility. At present, if libraries want to efficiently broaden the scope of public cultural services, they should introduce AR virtual reality technology as soon as possible, show the service model to readers in a brand new form, and use high-quality resource construction to burst out the library's own advantages.

*2.1. Construction of the Technical Hierarchical Model of Smart Library.* The library uses its own obvious resource advantages and integrates the advanced service performance of smart libraries, combines the library models built by predecessors, and introduces smart modern technology based on the theory of integration and collaboration to build a technical-level technical-service model that meets user needs. As shown in Figure 1, the current situation that libraries use a general service mode to meet user needs is broken, and the stable development of the library's service concept of "serving users and satisfying users" is promoted.

The architecture shown in Figure 1 is mainly composed of four parts: smart perception layer, network transmission layer, data resource layer, and smart application layer. The smart perception layer in the smart library is mainly composed of wearable devices, sensors, storage devices, RFID, and video. It consists of monitoring equipment and network monitoring equipment. Users visit the library under the information demand target. The intelligent equipment in the intelligent perception layer perceives, screens, and extracts the data traces generated during the reader's enjoyment of the library's application mode and services. Through the network transmission layer, according to the distance between the geographic location of the sensing device and the library collection data, the large amount of data collected by the intelligent sensing layer can be safely, efficiently, and quickly transmitted through the wireless network, triple play technology, and computer communication network.

The data resource layer is at the center of the technical model. It consists of data warehousing, data mining, cloud computing, information push, and semantic analysis technologies. It is mainly responsible for user data storage and format conversion, user data resource mining and calculation, and user personalized information needs. Task functions include prediction, recommendation, and management. The smart application layer relies on the data analysis provided by the data resource layer and is mainly constructed by virtual reality technology, multimedia, data visualization, and other technologies to realize library scene-based recommendation services, user personalized services, virtual reality services, and multimedia services. The improvement of smart space services and visualization services

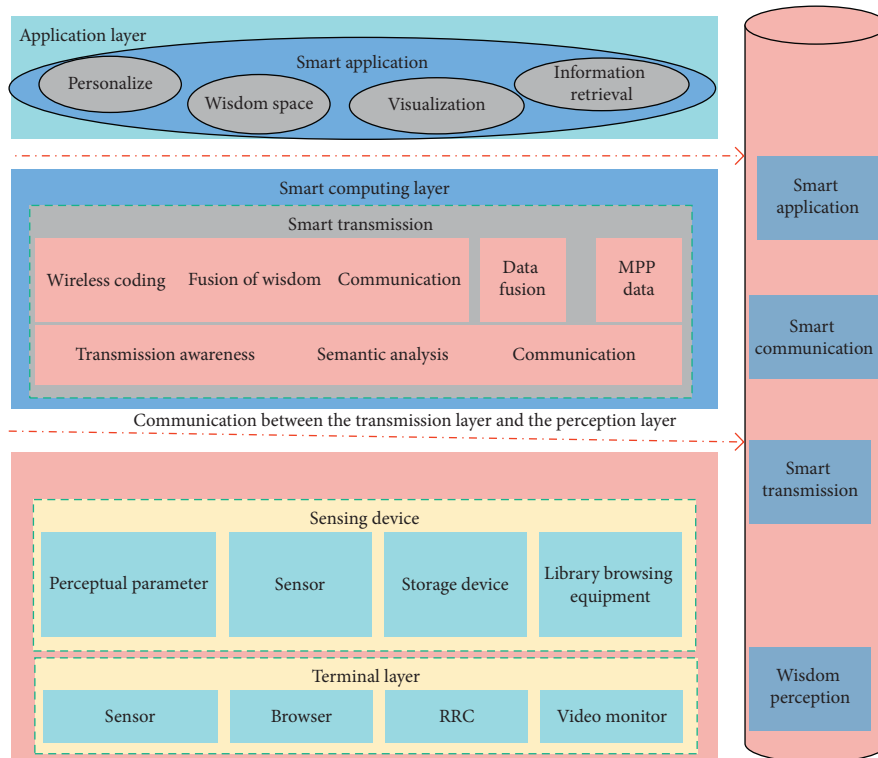


FIGURE 1: Architecture of the intelligent library information retrieval technology model construction system.

and the innovative development of smart libraries have realized the service process from sensing information to digging information, processing information, and finally discovering wisdom.

Virtual reality technology has computer-generated three-dimensional effects that integrate visual, tactile, and olfactory functions, allowing users to enjoy interactive services immersively with realistic and visual scenes. The multiple senses, visibility, permeability, and immersion characteristics of virtual reality technology make it popular in the library field. At present, virtual reality technology is mainly used in the virtual official buildings of libraries, allowing users to “walk” among them, freely associate with the virtual space, and obtain a three-dimensional and realistic user experience. When users “walk into” the virtual space of the library, they can understand the overall layout of the library and can also obtain information consulting services and browse the operation mode of the library in the most direct way of expression, allowing them to understand at a glance and fully embody the superiority of the smart library. It also appropriately compensates for the one-sided and localized information obtained on the library website and increases the user’s affinity for the virtual reality of the smart library. The application of visualization technology in smart libraries can realize the service functions of explicit tacit knowledge, clarification of fuzzy knowledge, and concrete abstract knowledge. In the process of smart library services, the collection of knowledge and the organization of knowledge are inseparable from the service criteria of “knowledge visualization.” Therefore, the smart library must achieve the visual effect of vivid knowledge services. Data

visualization technology is conducive to processing intricate data relationships, transforming data dimensions into visual dimensions, and then mining the knowledge structure and development trends hidden in the data.

*2.2. Application of Virtual Reality Technology in Smart Library.* In the actual application process, the manifestation of virtual reality technology can be divided into three types according to the user’s “immersion” and “interaction.”

*2.2.1. Virtual Reality Technology Display Mode Based on Computer Screen Display.* The real-world image or video captured by the camera equipment is input to the computer, synthesized with the virtual scene generated by the computer graphics system, and output to the screen display. The user sees the final enhanced scene picture or video on the screen. This form of expression is simple, in the enlightenment stage of the concept of virtual reality technology, and most of the things shown to users are display attributes, unable to operate and interact, and cannot bring much immersion to users. The implementation scheme of the virtual reality technology system based on computer screen display is shown in Figure 2.

*2.2.2. Display Mode Based on Optical Perspective.* This type of virtual reality technology needs to use display devices that emphasize the user’s vision and tactility, mainly helmet and glasses-type displays, to enhance the visual immersion by being close to the user’s body. Early AR products produced

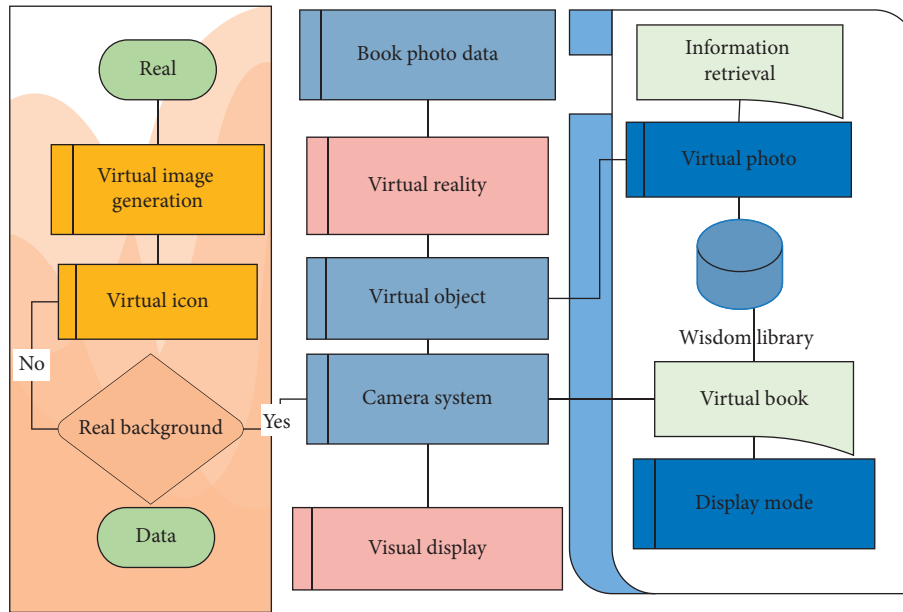


FIGURE 2: Operating principle model of virtual reality technology display mode based on computer screen display.

by electronic product companies such as Sony and Google are geared towards users in the form of helmets and glasses. Different from the computer screen display, the user can directly see the external real world. In fact, it uses a transparent optical synthesizer to project the signal of this virtual image to the eyes, and then because the real world is clearly visible, there is no way through shooting processing and artificial display. Its operating system is shown in Figure 3.

User experience design includes information construction, user interface design, human factor configuration design, and user experience evaluation. In order to complete these designs, they must be carried out in stages according to the requirements of user experience design, namely, the discovery of information needs based on user experience, the construction of user-oriented services based on resource integration, and feedback control based on information integration and service integration. These three aspects can be subdivided, as shown in Table 1.

The integration of information resources based on user experience directly faces users, highlights needs and services, and allocates limited resources to the service businesses that users care about most to ensure the efficiency of resource utilization. To do this, you need to analyze the information needs of users. The discovery process is to understand and analyze users' information needs from multiple angles. The measures taken include analyzing the current environment, understanding the real needs of users through surveys or interviews, deep mining and using the acquired user information, and coordinating user needs with the service organization's strategy and environmental requirements and then enter the design phase of user experience construction.

Construction design is the main part of user experience design, including information architecture, user interface design, human factor research, and user experience evaluation. Information architecture is a high-level information

design that focuses on the organization and presentation of information, and its purpose is to provide users with clear and understandable information. User interface design requires reasonable arrangement of interface elements on the basis of information construction, distinguishing the importance of information, expressing information in an easy-to-understand manner, and enabling users to interact with system functions. Human factor research and user experience evaluation are generally combined. Their responsibilities include user experience testing, researching user development, communicating with users, and passing these results to designers. The feedback control design requires the specification and description of the design process and user-oriented business, listening to opinions in many aspects, and designing a user-oriented integrated service feedback control system to ultimately improve the user experience design.

*2.3. Virtual Reality Technology Improves the Effectiveness of Information Retrieval Services and Management of Smart Libraries.* As the current popularity of virtual reality technology is not very high and it is not currently included in the development plan of the library business, the survey of this technology in the questionnaire must first start with the cognitive level of the surveyed, and it is easy to be surveyed. The questionnaire was designed based on the expression of the reader's understanding, and the cognitive results are shown in Table 2.

The survey results showed that 35 people did not understand the concept of virtual reality technology at all, accounting for 36.55% of the total number of people in the survey. Because the cognitive problem of the survey is also an important reference data, I do not think that choosing this option will lead to a whole answer. This questionnaire is invalid, which also reflects the current state of understanding of this technology in the industry to a certain

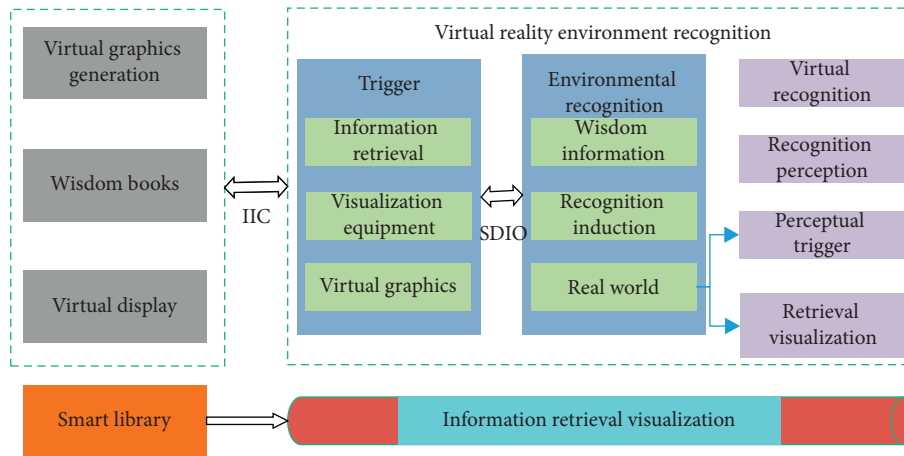


FIGURE 3: Display mode based on optical perspective.

TABLE 1: User experience design process.

| User needs                 | Experience building            | Feedback control    |
|----------------------------|--------------------------------|---------------------|
| Background analysis        | Concept + first prototype      | Design description  |
| User interview             | Site map                       | Process feedback    |
| User role and plot setting | Interactive model              | Performance control |
| Brainstorming              | Idea refinement                |                     |
| Group design practice      | User interface + visual design |                     |
|                            | User experience evaluation     |                     |

TABLE 2: User perception statistics.

| Variable  | Types of                 | Quantity | Percentage |
|-----------|--------------------------|----------|------------|
| Awareness | Don't understand         | 35       | 36.55      |
|           | Scan the QR code         | 34       | 38.71      |
|           | Sweep the reality object | 34       | 35.47      |
|           | Electronic games         | 28       | 31.17      |
|           | For life                 | 35       | 36.58      |

extent, and even if you do not understand the concept of virtual reality technology, you can express the expectations of the surveyed in the subsequent survey options. At the same time, I set the option of "scan QR code" in the research options. In fact, although the method of using QR code is similar to virtual reality technology, the two are not in a mutually compatible relationship. Codes do not belong to virtual reality technology. In the survey results, 14 people chose the QR code option alone, accounting for 14% of the total number of people. It can be understood that these users do use QR codes but they are still familiar with virtual reality technology. If combined with users who do not know anything about it, the proportion of respondents who have a weak awareness of virtual reality technology reaches 50.6%, which accounts for about half. This ratio reflects the current use of virtual reality technology in books. Among the remaining options, "scanning reality" is one of the main features of virtual reality technology, "electronic games" is a hot development field of current virtual reality technology, and "available for life services" means that all types of technologies including virtual reality technology are available. The selected development goals and those that

chose these options also accounted for the other half of all surveyed objects, reflecting the controversy of virtual reality technology in libraries.

The combination of virtual reality technology and library management and services has not only obtained theoretical support in the previous chapters but also should find the advantages of the combination of the two in real applications to enrich theoretical research and lay a foundation for popularization of practice. This section summarizes specific advantages by combining the results of the questionnaire survey and the author's reasonable assumptions. The survey results are shown in Table 3.

In the questionnaire survey on the advantages of virtual reality technology in library applications, 83 people think that virtual reality technology can provide more comprehensive services for library readers, accounting for 90.31%, and 73 people think it will enhance the interaction between the library and readers' sex, accounting for 77.43%, 57 people think that this will make paper books break the limit of paper media, accounting for 60.23%, 50 people think that virtual reality technology in the library will help library management, accounting for 55.9%, and 5 people think the advantage is not obvious, accounting for 4.4%. From the analysis of the above survey results, it can be seen that service is the primary consideration in the application of virtual reality technology in smart libraries, and its advantages can be better reflected in services. There is a certain gap between the research objects in management and the selection of services. This also shows that the proportion of virtual reality technology in the development of smart library services and management may also fit the selection of this survey.

TABLE 3: User advantage selection.

| Variable       | Types                            | Quantity | Percentage |
|----------------|----------------------------------|----------|------------|
| Advantage type | Full service                     | 83       | 90.31      |
|                | Improve interactivity            | 73       | 77.43      |
|                | Breakthrough paper media         | 57       | 60.23      |
|                | Improve management effectiveness | 50       | 55.9       |
|                | The advantage is not obvious     | 5        | 4.4        |

The number of users of smart libraries is large, and the level of information retrieval service is at the forefront of the times. It is the basis for smart libraries to adopt and popularize new technologies, and it is the main advantage. Compared with the more expensive and fixed-location navigation machine in the library, the virtual reality technology can be realized on the user's mobile phone, and the service can be provided through the design of the program, and the user's use method is flexible, easy to learn, and fast. As far as the development of smart libraries is concerned, the introduction of intelligent information services such as virtual reality technology improves service efficiency and implements the reader-centered concept. In the questionnaire survey, more than 75% of the respondents believe that virtual reality technology has improved the interaction between the library and the reader, and the smart library has the interaction between the reader and the smart library in each part. The application of virtual reality technology will make the way of interaction more intelligent. In terms of social value, if the use of virtual reality technology is promoted by the smart library of colleges and universities and the concept of new technology is popularized among young users, it will make college students who are good at using computers and mobile phones take advantage of this technology. The strong impression will also affect their future scientific and technological concepts and thus they will have a sense of dependence on the smart library and become loyal users. The group of young people is a group with a rapid popularity and a large number. It can also achieve certain effects in public libraries with evenly distributed age groups. Providing virtual reality technology services for older and traditional library user groups can achieve the effect of narrowing the social technological gap and completing large-scale technical literacy.

Any new technology may be a double-edged sword in the initial stage of application. Through the understanding of the functional characteristics of virtual reality technology, combined with the learning and understanding of the library, and the questionnaire survey and analysis, it is also found that the following points may limit its development, and user survey results are shown in Table 4.

In the questionnaire survey on the limitations of the application of virtual reality technology in libraries, 61 people think that the equipment of virtual reality technology is more expensive, accounting for 64.53%, 44 people think that the utilization rate of virtual reality technology in libraries is low, accounting for 48.38%, 28 people think that virtual reality technology is difficult to popularize as a new thing, accounting for 31.17%, 25 people think that virtual reality technology is complicated and difficult to use,

accounting for 25.82%, and 5 people think there are other reasons, accounting for 4.5%. Therefore, it can be seen from the survey results that the cost issue is the main issue in the application of virtual reality technology in the library. This is not only in the aspect of virtual reality technology. The library's funding is based on the overall consideration of each part of the library, so virtual reality of the cost of technical equipment should be carefully considered. Without experiments, the utilization rate is also worthy of attention. The low utilization rate of advanced and expensive equipment after the introduction will cause great losses to the library. The limitations are discussed in detail below.

### 3. Knowledge Service Extension and Information Retrieval Visualization of Smart Library

As a valuable and high-quality high-level knowledge service product, smart service is a perceptible, calculable, and visualized creative service, which will drive the technological upgrading, conceptual innovation, management reform, and service transformation of smart libraries. The library becomes an incubator for technological innovation and creativity and promotes the burst of imagination and creativity of users.

*3.1. Characteristics of Knowledge Service Extension of Smart Library.* The ultimate goal of the extension of the knowledge service of the smart library is to improve the knowledge service capability and level of the smart library and meet the increasingly diverse and individual needs of users. Specifically, the extension of the knowledge service of the smart library is to use the Internet of Things technology to realize the digitization of resources, the application of metadata harvesting, and the establishment of a data warehouse; as well as the storage and calculation of big data on the basis of data interconnection, forming a first-hand reliable information resources; on the basis of information collection, resources are reorganized, a knowledge base system through resource reengineering is built, and the knowledge of information is realized; a precise service platform based on situational awareness is built, and personalized knowledge service products is provided; the intelligence of knowledge is realized, and finally big data analysis tools are used to perform machine learning, mine user preferences, recommend personalized knowledge products, and achieve precise services. The goal of the extension of the knowledge service of the smart library is shown in Figure 4.



TABLE 4: Users' views on limitations.

| Variable             | Types                                  | Quantity | Percentage (%) |
|----------------------|--|----------|----------------|
| Types of limitations | Expensive                              | 61       | 64.53          |
|                      | Low utilization                        | 44       | 48.38          |
|                      | New things are difficult to popularize | 28       | 31.17          |
|                      | Complex operation                      | 25       | 25.82          |
|                      | Other reasons                          | 5        | 4.5            |

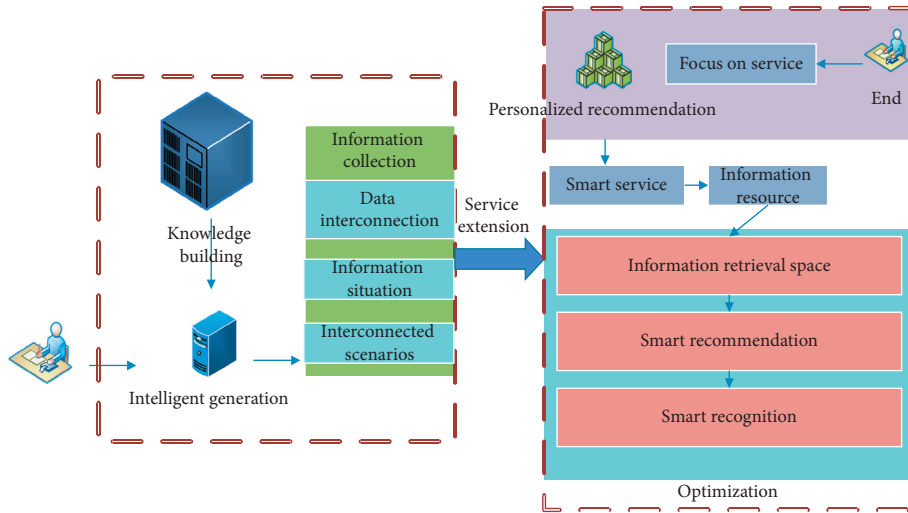


FIGURE 4: Goal-oriented diagram of the knowledge service extension of the smart library.

As shown in Figure 4, the interconnected context is the technical guarantee for the knowledge service context function of the smart library. The accessibility and popularization of the interconnected context depend on the supportive context technology of the Internet of Things and the ease of use and audience perception of the knowledge service platform. Resource context is the material basis of knowledge services in smart libraries. Resource reorganization and resource reengineering to improve the quality of resource content are related to the smooth development of user knowledge mining, knowledge association, knowledge utilization, and knowledge creation. In particular, human resources have become the first resource for the extension of knowledge services in smart libraries. It is the key to applying emerging technologies in the process of knowledge services and carrying out resource reengineering, knowledge creation, and smart services. The service context is the integration of the elements of the knowledge service field of Unicom's smart library under the combined effect of the technology interconnection context and the resource context. The personalized service and precise adaptation incentive function reflect the final performance of the knowledge service of the smart library. The interconnected context, resource context, and service context are cross-integrated and interacted in the knowledge service extension mechanism of the smart library, which together influence and determine the knowledge service level and users' perceived experience and satisfaction.

*3.2. Endogenous Power Mechanism Model of Knowledge Service Extension of Smart Library.* Based on the mechanism of interaction between the user and the smart library knowledge service system and the influence of the interaction function of the smart library knowledge service system on user behavior, this model is constructed from the different context dimensions of the interaction process between the user and the smart library knowledge service system. Users are the most basic and active force in the extension of knowledge services. The function of the knowledge service system of the smart library and the influence mechanism of user behavior are the user context, which mainly includes user needs and user experience. The three knowledge service context factors of the smart library are interconnection context, resource context, and service context. The corresponding ones are the ease of use, usefulness, and motivation of the correlation between the knowledge service function of the smart library and the user information interaction behavior. They include the ease of use of the space-time system, the ease of use of the supporting system, the usefulness of the resource construction system, the usefulness of the resource reengineering system, the standardized incentives of the service system, and the personalized incentives of the service system. The extension of the knowledge service function of the smart library and the construction of the mechanism model of user behavior are shown in Figure 5.

It can be seen from Figure 5 that user experience interaction and user demand promotion are accompanied by

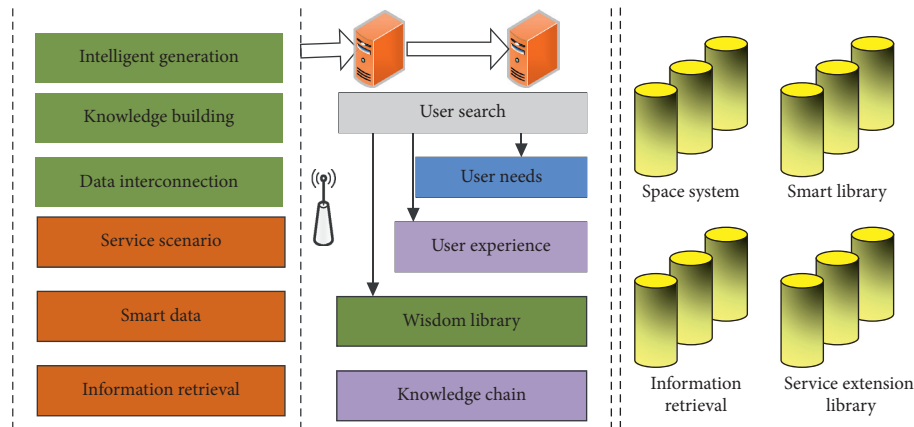


FIGURE 5: Mechanism model of endogenous power and user behavior of the extension of knowledge service in smart libraries.

both service experience characteristics and demand promotion functions. The interactive function of the knowledge service system of the smart library and the user's information interactive behavior work are through the following mechanism: the information of the knowledge service is presented to the user, and the user confirms the occurrence of the interactive behavior through self-perception. The psychological feelings of cognition, emotion, value, and so forth obtained by users through interactive behaviors are the result of comparison between user interaction perception and expectations, which directly affect the evaluation of knowledge service system functions and service incentive levels. In the process of information interaction, the user's impressive experience is accompanied by use and operation behavior, including the mastery of the knowledge service context function and the realization of needs; the user will always be based on the interactive experience and perception of the received customized information feedback and adjust their information interaction behavior. The functional elements of the knowledge service interconnection context, resource context, and service context of the smart library enhance and promote the data interconnection, knowledge construction, and wisdom generation of users, which have an impact on users' information interaction behavior, stimulate users' interactive interest at any time, and induce users' information interaction. The result of its interactive experience is changed. Users' knowledge needs to have new characteristics such as interconnection, sharing, knowledge integration, ubiquity, intelligence, and innovation. The user's interactive behavior in the knowledge service coincides with the ease of use of the knowledge service interconnection context, the usefulness of the resource context, and the motivation of the service context; that is, the user's psychological feelings and satisfaction with the knowledge service and reflecting a good experience of knowledge service quality is the whole process of the effect of the endogenous power of the knowledge service extension of the smart library.

The application of virtual reality technology in smart libraries has certain feasibility. First, the integration of the "three networks" and the construction of the Internet of Things provide networks and traditional equipment and

provide hardware support for the use of virtual technology. The integration of the "three networks" has achieved a high level of resource sharing, with higher communication speed, higher communication quality, and stronger communication security. As an important part of the new generation of information technology, the Internet of Things has realized the real connection of things with the help of the Internet. The development of the Internet of Things will greatly promote the development of smart libraries and VR technology. Second, the development of libraries requires the integration of virtual reality technology. Smart library is a new type of library facing the future and facing science and technology. It is an extension and expansion of traditional libraries. With the development of smart libraries, more and more information technology will be widely used in its function realization and service improvement. When VR technology has shown great advantages and potentials, its integration has improved the way of information services and enhanced the visibility and influence of smart libraries. Third, the characteristics of virtual reality technology are applicable to smart libraries. The immersion of virtual reality technology enables participants to exist in the virtual environment as subjects. The sense of interaction allows participants to get feedback from the virtual environment and realize interaction. Imagination will enable participants to expand more knowledge content through logical judgment, reasoning, and association based on the information they have obtained. It can be analyzed from the above three points that the application of virtual reality technology in smart libraries is very feasible.

### 3.3. Visualization of Information Retrieval in Smart Library.

For the design of the information visualization model of the smart library, it must not only conform to the basic system structure of the smart library but also fully meet the various requirements of information visualization. Therefore, for the design of the model, the following principles must be met:

- (1) Meeting the individual needs of users. The design of the model should be able to meet the needs of various users of the smart library, and different users have different preferences for the choice of visualization

methods. The personalization of user preferences is mainly in the results, that is, the way to provide personalized visualization results through the icon library.

- (2) Universality. The information resources contained in a smart library are massive, and there are many types of these information resources, including text, graphics, images, sounds, and videos. In the process of model design, it should be classified according to the information resources contained in the smart library to avoid the phenomenon that only information in a specific field is effective, so that the model can be used in a wider range of fields.
- (3) Convenience. Convenience here means that when the user visits the smart library, its visual interface is friendly and there is good interaction between the user and the system.
- (4) Interoperability. Information visualization technology faces the massive collection of information in smart libraries. This information is stored in various databases in different formats. The visualization system should be able to achieve undifferentiated access to information. In addition, the information contained in the smart library comes from different application fields, and the system should also implement interoperability between them.
- (5) The flexibility of the structure. The development prospects of visualization technology are unpredictable. Both information technology and computer technology are developing rapidly, and it is impossible for any technician to fully predict the future development.

Fuzzy C-means clustering algorithm is a typical information retrieval partitioning algorithm; its idea is to make the similarity between objects classified into the same category is the largest, and the similarity between different objects is the smallest. In the iterative optimization process, the FCM algorithm continuously updates the values of the various centers and the elements of the membership matrix until it approaches the minimum value of the following criterion function:

$$T_n(O, P) = \sum_{j=1}^N \sum_{i=1}^c o_{ij} w_{ij}^2. \quad (1)$$

Regarding feature extraction, in the test sample data set, there are a total of 988 keywords, and the number of keywords that are different from each other reaches 628.  $w$  is the weight of the number of keywords. After the data dimensionality reduction process, the remaining unique keywords are 113, which greatly reduces the aggregation. The data dimension of the class algorithm in the document space vector matrix  $R$  is stored in a text file as the data source of the FCM algorithm in Matlab. The number of iterations of running the FCM algorithm is 100, and the clustering result when the clustering objective function value is the smallest is taken out as the final result output. The results of the five experiments are shown in Table 5.

The clustering results of the third run in Table 5 are the best, and only 2 documents are misclassified in each of the three categories. In general, when the number of FCM documents is small, the clustering quality is better, with an average correct rate of 97.4%. But sometimes the result is not very stable, mainly due to the limitations of the FCM algorithm itself; that is, the random initialization of the center point has a greater impact on FCM.

Since keywords can directly indicate the subject of the literature and the characteristics of the subject, the similarity of the literature can be directly reflected by the similarity of the keywords. In order to cluster the documents, the document similarity matrix is first defined, and the similarity is expressed by Euclidean distance. The following definition is based on the following assumptions: suppose that the total number of documents to be clustered is  $n$ , the total number of different keywords in all documents is  $m$ , the keyword set is  $S$ , and the number of categories is  $k$ . The document similarity matrix ( $n * n$ ) is defined as

$$\begin{bmatrix} W_{11}, W_{12}, \dots, W_{1n} \\ W_{21}, W_{22}, \dots, W_{2n} \\ \dots \\ W_{n1}, W_{n2}, \dots, W_{nm} \end{bmatrix}. \quad (2)$$

The similarity of two keywords is defined as

$$W(\lambda_i, \lambda_j) = \begin{cases} \frac{\kappa}{\kappa_i + \kappa_j - \kappa}, & k \geq 4, \\ 0, & \kappa < 4. \end{cases} \quad (3)$$

The components of the literature keyword matrix are defined as

$$Q_{ij} = \max(W(\lambda_i, \lambda_j)). \quad (4)$$

The clustering objective function is defined as follows:

$$E = \frac{\sum_{j=1}^k \sum_{i=1}^m (y_i - y_j)^2}{m_j}. \quad (5)$$

The selection operation is adopted to select good individuals from the current group and decide which individuals can enter the next generation. First, the individuals are sorted according to the fitness function from large to small, and the first  $h$  individuals are copied as new individuals directly into the next generation, and the fitness of the remaining individuals is calculated as follows:

$$P(D) = \left[ c + (d - c) \frac{M - \text{rand}(D)/M - d}{(M - c)} \right]. \quad (6)$$

In order to test the feasibility and effectiveness of the information retrieval of the smart library, the value of 207 documents in the life is taken as the test data set. The genetic algorithm parameter is 50, the probability of mutation is represented by  $P$ , the value is 0.15, and the intersection value is 0.0002. The value is 0.76, and the maximum number of iterations is represented by  $T$ , with a value of 100. After 50

TABLE 5: FCM clustering results.

| Number of experiments | Information science (52) |       | Philology (71) |       | Library science (87) |       | Objective function |
|-----------------------|--------------------------|-------|----------------|-------|----------------------|-------|--------------------|
|                       | Correct                  | Error | Correct        | Error | Correct              | Error |                    |
| 1                     | 47                       | 2     | 68             | 1     | 84                   | 2     | 101.2245           |
| 2                     | 48                       | 4     | 71             | 2     | 83                   | 3     | 100.4342           |
| 3                     | 51                       | 2     | 69             | 2     | 84                   | 2     | 98.9119            |
| 4                     | 48                       | 3     | 67             | 2     | 83                   | 3     | 102.5346           |
| 5                     | 51                       | 2     | 67             | 3     | 84                   | 2     | 99.2325            |

experiments, the result of the average objective function is shown in Figure 6.

It can be seen from Figure 6 that, in the iterative process of the FCM algorithm, the objective function can converge quickly, and the value of the objective function is greater than that of the GA algorithm, indicating that the convergence accuracy of the FCM algorithm is inferior to the GA algorithm. GA algorithm converges slowly, but the average accuracy of classification can reach more than 99%, which is better than FCM algorithm. As the GA algorithm draws on the idea of genetics in biology, it searches for the optimal solution through repeated iterations of “survival of the fittest.” Therefore, the optimization ability of GA is better than the FCM clustering algorithm, but the disadvantage is that the GA calculation speed is slow and many parameters need manual intervention.

**3.4. Information Retrieval Visualization Construction.** For the design of the information visualization model of the smart library, it must not only conform to the basic system structure of the smart library but also fully meet the various requirements of information visualization. Therefore, for the design of the model, the following principles must be met:

- (1) Meeting the individual needs of users. The design of the model should be able to meet the needs of various users of the smart library, and different users have different preferences for the choice of visualization methods. The personalization of user preferences is mainly in the results, that is, the way to provide personalized visualization results through the icon library.
- (2) Universality. The information resources contained in a smart library are massive, and there are many types of these information resources, including text, graphics, images, sounds, and videos. In the process of model design, it should be classified according to the information resources contained in the smart library to avoid the phenomenon that only information in a specific field is effective, so that the model can be used in a wider range of fields.
- (3) Convenience. Convenience here means that when the user visits the smart library, its visual interface is friendly and there is good interaction between the user and the system.
- (4) Interoperability. Information visualization technology faces the massive collection of information in

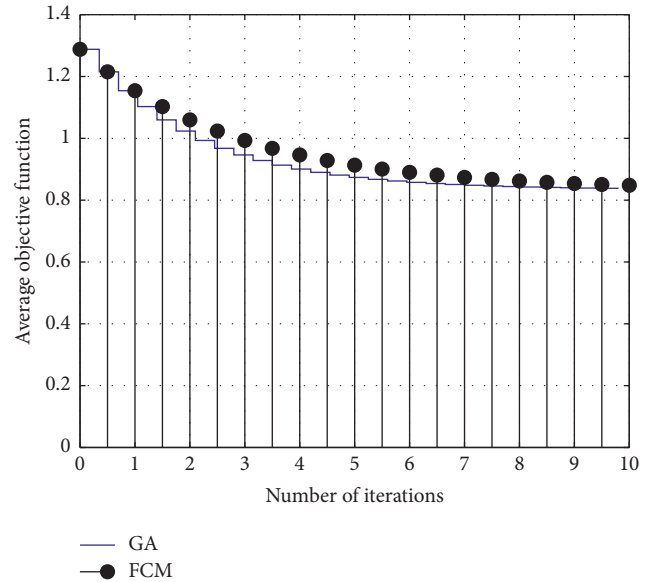


FIGURE 6: Comparison of the relationship between objective function and evolutionary algebra.

smart libraries. This information is stored in various databases in different formats. The visualization system should be able to achieve undifferentiated access to information. In addition, the information contained in the smart library comes from different application fields, and the system should also implement interoperability between them.

- (5) The flexibility of the structure. The development prospects of visualization technology are unpredictable. Both information technology and computer technology are developing rapidly, and it is impossible for any technician to fully predict the future development. Therefore, when designing a visualization system, full consideration should be given to the standardization of the database and the expansion of the system.

The information visualization model of the smart library designed in this paper is shown in Figure 7, which is based on the combination of the basic architecture of the smart library and the reference model of information visualization. As shown in Figure 7, the proposed smart library information visualization model includes a total of six modules: source data module, original database module, feature

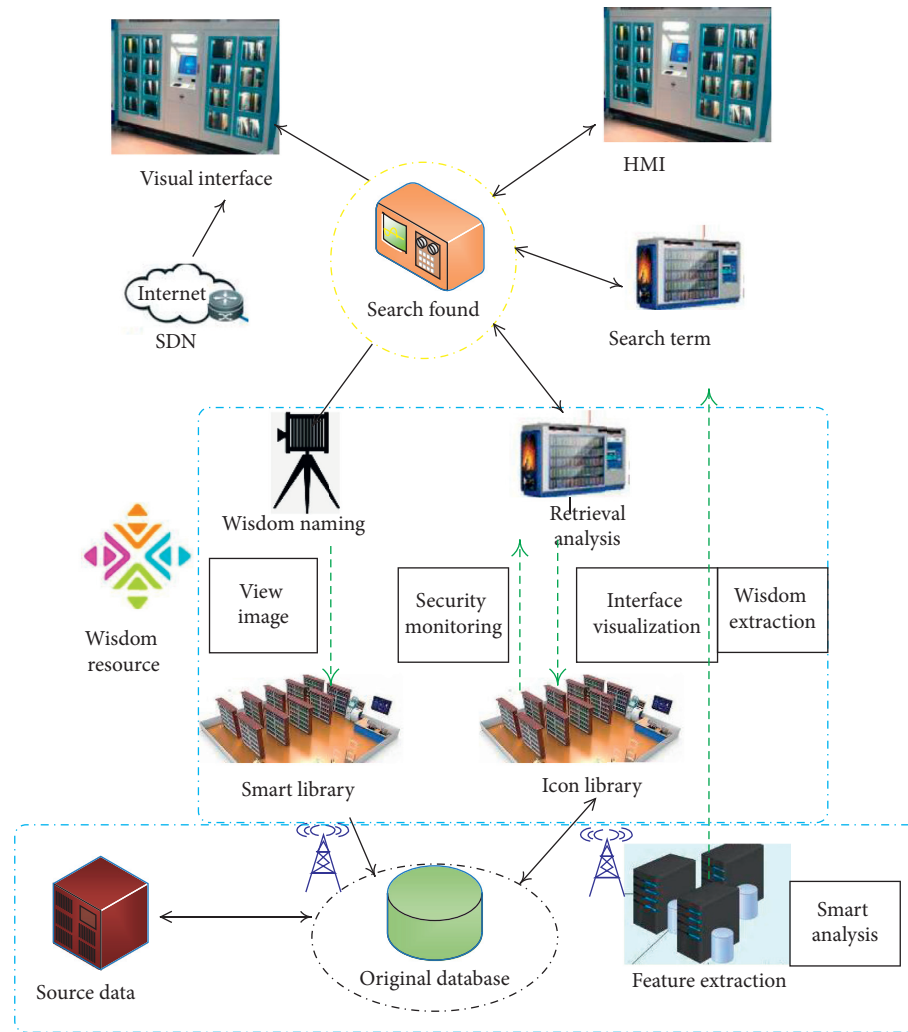


FIGURE 7: Information visualization model of smart library.

database module, view object module, visualization interface module, and extended function slot module.

Based on the above considerations, the L-Apriori algorithm is described as follows:

- (1) Classify the collection of books, combine the book classification method, and divide the collections into professional books and cross-professional books according to their professional background. Professional books for each profession are definitely different.
- (2) Preprocess the historical borrowing data, and set up subdatabases according to profession and year. Each subdatabase only contains the historical borrowing data of previous students in the major, which relates to the borrowing information of professional books and cross-professional books.
- (3) Data association rule mining of professional books: data mining is performed on subdatabases through Apriori algorithm; then the frequent item set A of each subdatabase must be the borrowing information of this professional book, and association rules can be extracted for recommendation; at the same time, get the frequent item set B of cross-professional books.
- (4) Prune the frequent itemsets obtained by Apriori algorithm mining, delete the frequent itemsets of professional books, and keep only the frequent itemsets of cross-specialties.
- (5) Compare the frequent itemsets A of the first subdatabase a and the frequent itemsets B of the second subdatabase b, find out the same parts, and put the same parts into the frequent itemsets C of the merged new data category c.
- (6) For frequent itemsets, scan b to obtain the support degree  $sup_x$  in b and  $sup_x$  plus the support degree  $sup$  in A; if the sum of the two is greater than or equal to the minimum support degree, put them in C. Similarly, for frequent itemsets, scan a and recalculate its support; and if it is greater than or equal to the minimum support, put it in C.
- (7) Repeat steps 5 and 6 until all the subdatabases are merged to form a new frequent itemset and

association rules, and the extracted association rules are recommended to users in a visual way.

#### 4. Experimental Verification

In order to verify the efficiency of the L-Apriori algorithm, a comparative test was conducted with the Apriori algorithm. The test data is the student borrowing data of a smart library from 2011 to 2012 in the last semester. The test results and the required time are observed through different mining algorithms. The required software and hardware environments are the same. The performance comparison of the two algorithms is shown in Figure 8.

Through the comparison and analysis of the test results and the required time, L-Apriori is more effective than April in the frequent itemset mining of library borrowing data, and the result is simpler, mainly because the L-Apriori algorithm pruned some of the frequent items of professional books in the process of subdatabase frequent itemset integration, and the pruned frequent items of professional books directly extracted the association rules and recommended them to users in the form of knowledge. The frequent items of cross-professional books are integrated one by one to form a frequent itemset of cross-professional books, which is much less than the frequent itemset obtained by April and naturally requires less time. Both the cross-professional book association rules obtained by the L-Apriori algorithm and the professional book association rules obtained by the Apriori algorithm can be recommended to users.

The recommendation service methods of the interviewed smart libraries that provide recommendation services can be divided into two types: personalized recommendation and nonpersonalized recommendation. As shown in Figure 9, about 56% of the interviewed smart libraries provide nonpersonalized recommendation services, such as “new book recommendation,” “borrowing ranking,” “hot review books,” “librarian recommendation,” and “reader recommendation.” Only about 23.5% of the provincial smart libraries have achieved personalized recommendations, but about 72% of them need to rely on third-party search systems to complete personalized search recommendations. Personalized search recommendations mainly rely on user search content to make recommendations, with similar recommendation forms and relatively single content, such as “Guess you like,” “Related Borrowing,” “Related Collection,” “Other Works by Authors of the Same Name,” and “Borrowing Relationship Diagram.” The survey found that the third-party systems that provide personalized search recommendations are mainly the ILAIII knowledge portal search system and the Interlib system. The smart library has completed personalized recommendation under its own recommendation system. When a user logs in to the service platform of the smart library, the recommendation system can analyze the user’s interest and preferences based on the user’s historical data, provide personalized recommendations of “Guess you like” on the homepage, and provide nonpersonalized recommendations of “everyone cares” or dynamic recommendation.

Using the information retrieval visualization API, event processing of the visualization interface can be performed, and interactive control can be added. This means that multiple visualization views can be coordinated, and data flow and control flow can be managed with the server through event processing. Information retrieval visualization provides more than 20 chart types, including the chart type that comes with the API and many visualization chart types developed by third parties, such as tag clouds. Therefore, with the increase of the API applications, the optional visualization types will gradually increase. As shown in Figure 10, the view elements are manipulated by responding to mouse click events. The information retrieval visualization API can also call the information retrieval visualization API to update the visualization view through asynchronous interaction with the server without updating the entire page.

However, from the perspective of the types of visualization, there is a lack of many classic views in the field of information visualization, such as graph-based visualization types and many tree-based visualization views (e.g., hyperbolic trees, radial trees, etc.). It lacks focus + context and overview + detail visualization.

The implementation of questionnaire survey mainly includes two parts: questionnaire survey and individual interview. All the users who participated in the questionnaires selected in this article are those who have used mobile virtual reality technology and are between 20 and 40 years of age. The author first explained the basic concepts of mobile virtual reality technology and then briefly introduced the mobile virtual reality technology prototype system and its functions designed and implemented. Finally, the software was installed on the mobile phone with the Android operating system and the subjects personnel conduct operation and experience. A total of 55 questionnaires were distributed in the survey and 55 were returned, of which 48 were valid questionnaires, with an effective rate of 95%. After collecting the questionnaires, the author randomly selected several questionnaires and conducted individual interviews with the corresponding subjects to discuss related issues that need to be understood.

It can be seen from Figures 11 and 12 that the four categories of user experience scores are not high, the highest score is behavioral experience, with 55 points, and users are basically satisfied. Since the designed mobile virtual reality technology prototype does not involve a social module, users will basically not have a social experience, so the social experience score is low and users are very dissatisfied.

In order to verify the clustering performance of the information retrieval visualization algorithm, the International Standard (IRIS) classification data is used for testing. The data set uses the characteristics of the Orioles as the data source. The data set contains 150 data sets, divided into 3 categories, each with 50 data, and each data contains 4 attributes. It is very commonly used in data mining and data classification for test set and training set. In the PSO algorithm, the inertia weight and learning coefficient adopt linear change strategy. In the information retrieval visualization algorithm, the number of IS-PSO iterations is 20, and

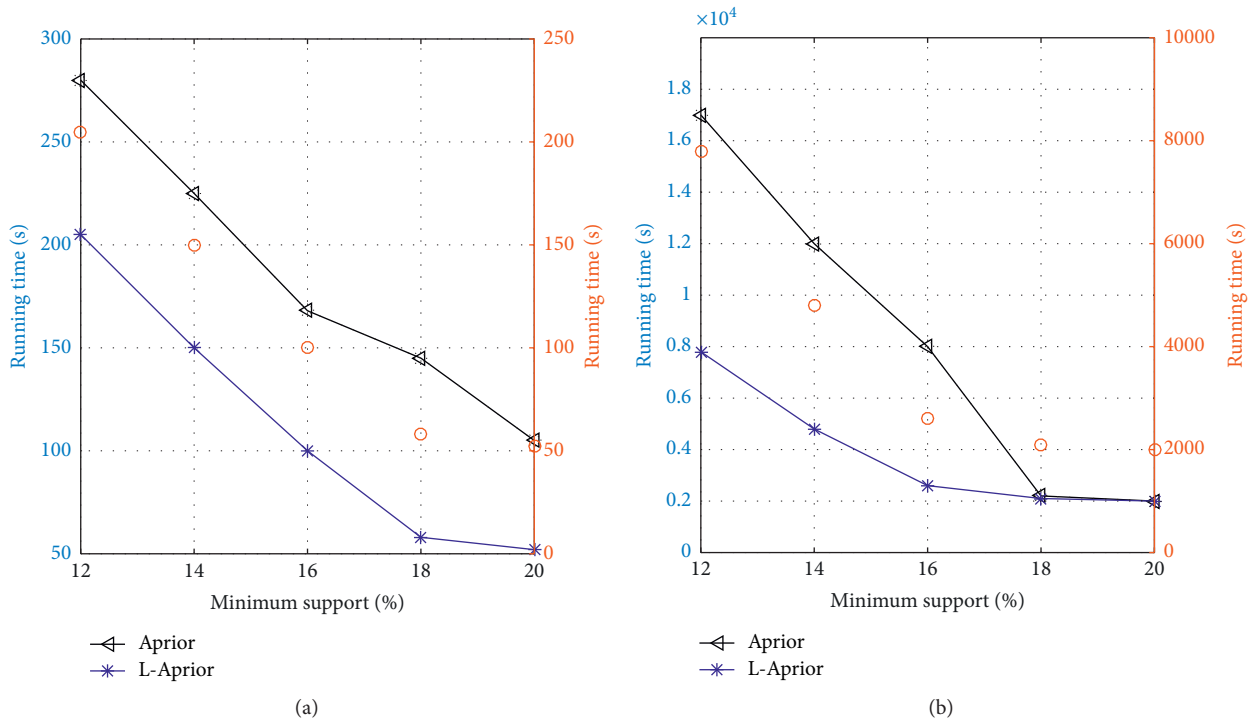


FIGURE 8: Performance analysis of L-Apriori algorithm.

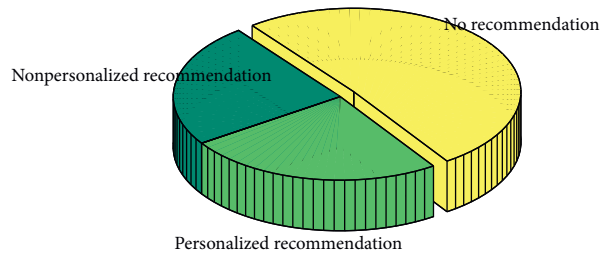


FIGURE 9: Specific content of the recommendation service provided by the smart library.

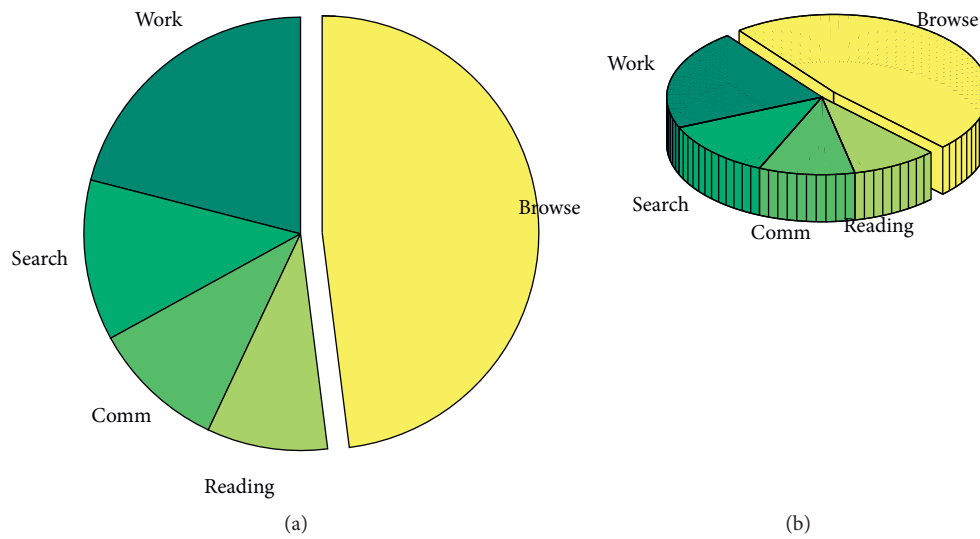


FIGURE 10: Pie chart (view events responding to mouse click).

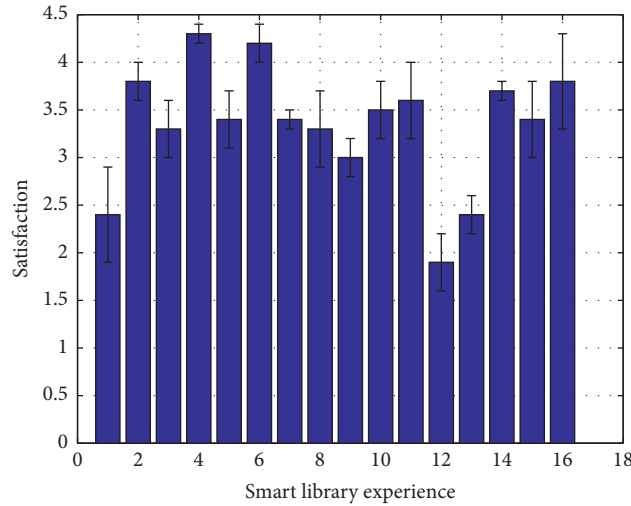


FIGURE 11: Histogram of satisfaction with various elements of user experience.

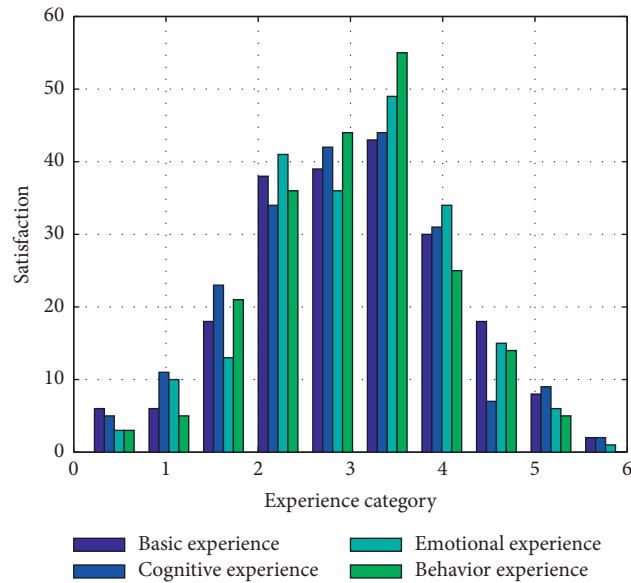


FIGURE 12: Histogram of user experience category satisfaction.

the number of FCM iterations is 80. The results are compared with the results of IS-PSO algorithm and FCM algorithm at 100 iterations. The results of multiple tests of the two algorithms on the same data set are relatively stable. Table 6 shows the comparison of the average performance of the three algorithms.

From the perspective of the correct rate of classification, the correct rate of FCM algorithm classification is only 88.68%, while IS-PSO and information retrieval visualization algorithms have reached a higher accuracy rate; from the perspective of running time, IS-PSO runs the most due to the complexity of the algorithm. The shortest running time of the FCM algorithm is more than 17 seconds, while the running time of the information retrieval visualization algorithm is in the middle. From the objective function, the

information retrieval visualization algorithm is the basis for finding a better center point in the PSO algorithm. On the other hand, with the advantage of fast convergence of FCM, the optimization accuracy is the best, while the IS-PSO algorithm could not exceed the information retrieval visualization algorithm in the optimization accuracy due to the limitation of running algebra.

Figure 13 shows a simulation diagram of the objective function optimization of the two algorithms. The FCM algorithm has the fastest convergence speed but is premature. The IS-PSO algorithm has the slowest convergence speed and the worst optimization performance within the range of 100 iterations. The information retrieval visualization algorithm is consistent with the IS-PSO algorithm in the first 20 generations, but in 20 after the first generation, due to the



TABLE 6: Comparison of average performance of algorithms.

| Algorithm             | Correct rate (%) | Operation hours | Objective function |
|-----------------------|------------------|-----------------|--------------------|
| FCM algorithm         | 88.68            | 0.138668        | 0.0515             |
| IS-PSO                | 94.66            | 18.358748       | 0.0485             |
| Information retrieval | 96.2             | 3.964959        | 0.0384             |

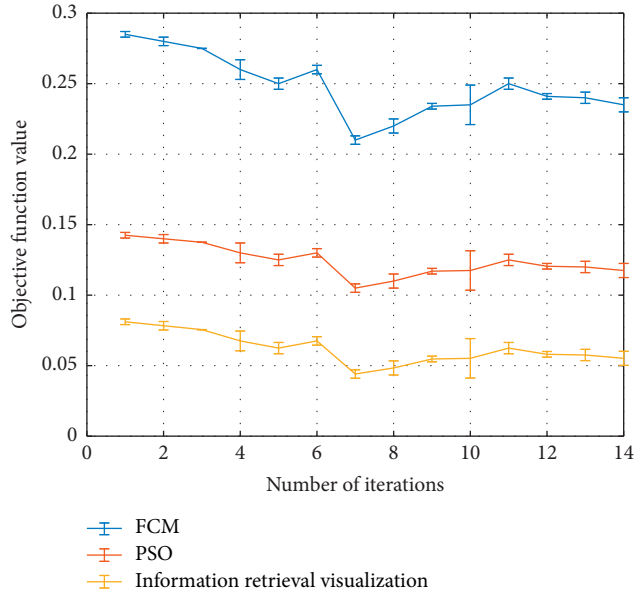


FIGURE 13: Simulation diagram of objective function.

FCM algorithm, the convergence speed is accelerated, and a better optimization effect can be quickly achieved.

As shown in Figure 13, although the efficiency of information retrieval visualization algorithm is slower than FCM, its optimization performance is better than FCM. The information retrieval visualization algorithm combines the advantages of higher optimization accuracy of PSO and the characteristics of fast convergence of FCM, so the information retrieval visualization algorithm can be better applied to document clustering in smart libraries.

## 5. Conclusion

Inner motivation is the decisive force to promote the increase of contextual functions of new knowledge services. The key to exploring the extension and sustainable development of knowledge services is to explore the intrinsic motivation of new knowledge service enhancements. Therefore, it is necessary to comprehensively consider the development and changes of user experience and perception in the knowledge service process, continuously improve the user's interactive experience and perception, continuously improve the new knowledge service context function of the smart library, and comprehensively and evenly improve the new knowledge service context function design and development to better meet the multidimensional applications and individual needs of users. External motivation is the driving force and supporting force to promote the contextual function of new knowledge

services, which mainly comes from the improvement of resource content quality and basic conditions. Resource context is an important force that determines the function of the new knowledge service, and the maximization of resource content attributes and value is the basic force to promote the interactive function of the new knowledge service context. According to the internal mechanism of knowledge service extension of smart library, the internal power and external power are actively and effectively used to form a dynamic and balanced situational state in order to promote the improvement of the overall function of the library's new knowledge service. By building a smart technology system of smart perception layer, network transmission layer, data resource layer, and smart application layer, with the help of various sensing devices in the perception layer, network technology in the network transmission layer, data mining and cloud computing in the data resource layer technology, virtual reality and augmented reality technology at the smart application layer complete the collection, distribution, and organization of user data and collection data and realize user personalized recommendation services, scene-based services, multimedia services, smart space services, visualization services, and virtual realistic service. However, smart libraries also need diversified smart countermeasures to solve various problems in the construction process and establish an interactive smart platform for the multifunctional smart library system to check for deficiencies and make up for the lack of public welfare services. Under the banner of advocacy, we can focus on user needs and implement the service standards of "precise demand and high-quality services" to create a smart space, to realize the interconnection of physical space, to increase the comfort and humanistic care of the space, and to embody the concept of a smart library for the convenience of readers and green development. In addition, the integration of online and offline libraries, libraries and bookstores, libraries and logistics companies, integrated service solutions, and other integrations are also being further explored, and integration is becoming the main form of the development of smart libraries.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The author declares that there are no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- [1] J. Hu, D. Wang, and B. Zhang, "Research on library information visualization retrieval technology based on readers' needs," *IOP Conference Series: Materials Science and Engineering*, vol. 569, pp. 52049–52054, 2019.
- [2] X. Y. Wei, Y. Wang, X. P. Yan, Y. F. Tian, and B. Wu, "Design and realization of the 3D electronic chart based on GIS and virtual reality technology," *Applied Mechanics and Materials*, vol. 744–746, no. 746, pp. 1669–1673, 2015.
- [3] A. Rosa-Pujazón, I. Barbancho, L. J. Tardón, and A. M. Barbancho, "A virtual reality drumkit simulator system with a kinect device," *International Journal of Creative Interfaces and Computer Graphics*, vol. 6, no. 1, pp. 72–86, 2015.
- [4] M. Cook, "Virtual serendipity: preserving embodied browsing activity in the 21st century research library," *The Journal of Academic Librarianship*, vol. 44, no. 1, pp. 145–149, 2018.
- [5] Y.-C. Chen and C.-H. Lin, "Image-based airborne lidar point cloud encoding for 3D building model retrieval," *ISPRS—International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLI-B8, pp. 1237–1242, 2016.
- [6] J. B. Julian, J. Ryan, R. H. Hamilton, and R. A. Epstein, "The occipital place area is causally involved in representing environmental boundaries during navigation," *Current Biology*, vol. 26, no. 8, pp. 1104–1109, 2016.
- [7] H. R. Dimsdale-Zucker, M. Ritchey, and A. D. Ekstrom, "CA1 and CA3 differentially support spontaneous retrieval of episodic contexts within human hippocampal subfields," *Nature Communications*, vol. 9, no. 1, pp. 294–303, 2018.
- [8] L. A. Liikkanen and P. Åman, "Shuffling services: current trends in interacting with digital music," *Interacting with Computers*, vol. 28, no. 3, pp. 352–371, 2016.
- [9] C. H. Lee, J. Ryu, S. H. Lee, H. Kim, and I. Lee, "Functional cross-hemispheric shift between object-place paired associate memory and spatial memory in the human hippocampus," *Hippocampus*, vol. 26, no. 8, pp. 1061–1077, 2016.
- [10] Z. Zhang, M. Xu, and J. Huang, "Proposals to promote the development of virtual reality in China-based on patent econometric analysis," *Engineering*, vol. 10, no. 5, pp. 291–304, 2018.
- [11] W. Zhou, J. Jia, and C. Huang, "Web3D learning framework for 3D shape retrieval based on hybrid convolutional neural networks," *Tsinghua Ence and Technology*, vol. 25, no. 1, pp. 93–102, 2019.
- [12] U. Dangel, Q. Bragard, P. Mcdonagh, A. Ventresque, and L. Murphy, "Can road traffic volume information improve partitioning for distributed SUMO?" *Modeling Mobility with Open Data*, vol. 13, pp. 61–74, 2015.
- [13] I. Bartolini, V. Moscato, R. G. Pensa et al., "Recommending multimedia visiting paths in cultural heritage applications," *Multimedia Tools and Applications*, vol. 75, no. 7, pp. 3813–3842, 2016.
- [14] A. Liu, W. Li, W. Nie, and Y. Su, "3D models retrieval algorithm based on multimodal data," *Neurocomputing*, vol. 259, no. 11, pp. 176–182, 2017.
- [15] H. J. Spiers and S. J. Gilbert, "Solving the detour problem in navigation: a model of prefrontal and hippocampal interactions," *Frontiers in Human Neuroence*, vol. 9, pp. 125–138, 2015.
- [16] S. Serino and G. Riva, "How different spatial representations interact in virtual environments: the role of mental frame syncing," *Cognitive Processing*, vol. 16, no. 2, pp. 191–201, 2015.
- [17] I. Lovrek, R. J. Howlett, and L. C. Jain, *Knowledge-based Intelligent Information and Engineering Systems*, vol. 28, Springer, Berlin, Germany, 2016.
- [18] G. Margetis, X. Zabulis, S. Ntoa et al., "Enhancing education through natural interaction with physical paper," *Universal Access in the Information Society*, vol. 14, no. 3, pp. 427–447, 2015.
- [19] D.-L. Dinh, S. Lee, and T.-S. Kim, "Hand number gesture recognition using recognized hand parts in depth images," *Multimedia Tools and Applications*, vol. 75, no. 2, pp. 1333–1348, 2016.
- [20] K. Kwon, H. Park, and S. Jung, "Dynamic scheduling method for cooperative resource sharing in mobile cloud computing environments," *KSII Transactions on Internet & Information Systems*, vol. 10, no. 2, pp. 484–503, 2016.
- [21] S. Serino, D. Mestre, P. Mallet, J.-M. Pergandi, P. Ciproso, and G. Riva, "Do not get lost in translation: the role of egocentric heading in spatial orientation," *Neuroscience Letters*, vol. 602, pp. 84–88, 2015.
- [22] F. Tian, X. Shen, X. Liu, and M. Cao, "Image tagging by semantic neighbor learning using user-contributed social image datasets," *Tsinghua Science and Technology*, vol. 22, no. 6, pp. 551–563, 2017.
- [23] Y. Wang, J. Yue, Y. Dong, and Z. Hu, "Review on kernel based target tracking for autonomous driving," *Journal of Information Processing*, vol. 24, no. 1, pp. 49–63, 2016.
- [24] G.-G. Du, C.-L. Yin, M.-Q. Zhou et al., "Isometric 3D shape partial matching using GD-DNA," *Journal of Computer Science and Technology*, vol. 33, no. 6, pp. 1178–1191, 2018.
- [25] Y. Kassahun, B. Yu, A. T. Tibebe et al., "Surgical robotics beyond enhanced dexterity instrumentation: a survey of machine learning techniques and their role in intelligent and autonomous surgical actions," *International Journal of Computer Assisted Radiology and Surgery*, vol. 11, no. 4, pp. 553–568, 2016.
- [26] L. Su and F. Lu, "Visual facial expression modeling and early predicting from 3D data via subtle feature enhancing," *Multimedia Tools and Applications*, vol. 75, no. 20, pp. 12563–12580, 2016.
- [27] Yo-S. Ho and H. J. Kim, "Advances in multimedia information processing—PCM 2004," *Lecture Notes in Computer Ence*, vol. 5879, no. 7499, pp. 201–204, 2015.
- [28] L. Hou, X. Wang, and M. Truijens, "Using augmented reality to facilitate piping assembly: an experiment-based evaluation," *Journal of Computing in Civil Engineering*, vol. 29, no. 1, 2015.
- [29] G. Qiu, K. M. Lam, and K. Hitoshi, "Advances in multimedia information processing—PCM 2010," *Lecture Notes in Computer Ence*, vol. 5879, no. 7499, pp. 201–204, 2015.
- [30] M. Koeva, M. Luleva, and P. Maldjanski, "Integrating spherical panoramas and maps for visualization of cultural heritage objects using virtual reality technology," *Sensors*, vol. 17, no. 4, pp. 829–841, 2017.
- [31] S. Romano, N. Capece, U. Erra, G. Scanniello, and M. Lanza, "On the use of virtual reality in software visualization: the case of the city metaphor," *Information and Software Technology*, vol. 114, pp. 92–106, 2019.
- [32] S. S. Yeşilyaprak, M. Ş. Yıldırım, M. Tomruk, Ö. Ertekin, and Z. C. Algun, "Comparison of the effects of virtual reality-based balance exercises and conventional exercises on balance and fall risk in older adults living in nursing homes in Turkey," *Physiotherapy Practice*, vol. 32, pp. 191–201, 2015.

- [33] M. Kim, J. Lee, and C. Jeon, "A study on interaction of gaze pointer-based user interface in mobile virtual reality environment," *Symmetry*, vol. 9, no. 9, pp. 189–197, 2017.
- [34] L. Bloomberg, J. Meyers, and M. Braverman, "An open-access web-based medical image atlas for collaborative medical image sharing, processing, web semantic searching and analysis with uses in medical training," *Research and Second Opinion of Cases*, vol. 92, no. 4, pp. 195–202, 2018.
- [35] F. Xu, "Applied-information technology in reconstruction of crime scene based on virtual reality," *Advanced Materials Research*, vol. 1046, pp. 465–468, 2014.
- [36] H. X. Liu and Z. H. Teng, "The exhibition design and implementation based on virtual reality technology," *Advanced Materials Research*, vol. 998-999, no. 999, pp. 1270–1273, 2014.
- [37] P. Seeling, "Towards quality of experience determination for video in augmented reality settings," *Signal Processing Image Communication*, vol. 33, pp. 41–50, 2014.
- [38] K. Melemez, G. Di Gironimo, and G. Esposito, "Concept design in virtual reality of a forestry trailer using a QFD-TRIZ based approach," *Turkish Journal of Agriculture & Forestry*, vol. 37, no. 6, pp. 789–801, 2014.
- [39] X. He, C. Shao, and Y. Xiong, "A new similarity measure based on shape information for invariant with multiple distortions," *Neurocomputing*, vol. 129, pp. 556–569, 2014.
- [40] G. Qin and Z. Xiao-Xue, "Solution and visualization of two kinds of diffusion problems based on mathematica software," *Journal of Hub Normal University*, vol. 2, pp. 342–357, 2018.

## Research Article

# Landscape Image Layout Optimization Extraction Simulation of 3D Pastoral Complex under Big Data Analysis

Juan Du <sup>1,2</sup> and Yuelin Long<sup>2</sup>

<sup>1</sup>College of Horticulture, Hunan Agriculture University, Changsha 410000, Hunan, China

<sup>2</sup>College of Art, Hunan City University, Yiyang 413000, Hunan, China

Correspondence should be addressed to Juan Du; [dujuan@stu.hunau.edu.cn](mailto:dujuan@stu.hunau.edu.cn)

Received 16 October 2020; Revised 13 November 2020; Accepted 17 November 2020; Published 29 November 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Juan Du and Yuelin Long. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Big data has brought about opportunities for landscape architecture, changing the design thinking of layout optimization simulation, expanding the platform for public participation in layout optimization simulation design, reflecting social and humanistic care, and promoting the integration of discipline cooperation and data. At the same time, it also brings about challenges. The proposal of data theory, the acquisition and analysis of data, and the protection of privacy are all issues that we need to face and solve. First, build a layout optimization simulation program under the background of big data. Follow the procedures of conventional layout optimization simulation, add big data analysis technology to the preliminary analysis, layout optimization simulation design, and later evaluation management, introduce new data processing methods, and build a theoretical framework of landscape optimization simulation methods that integrate new and traditional data. Second, studying the evolution of the landscape image layout of the three-dimensional pastoral complex is different: the overall landscape image layout of the three-dimensional pastoral complex shows a trend of fragmentation and heterogeneity. Among them, the arable land is gradually fragmented, the wetland and water area are simplified in form, the woodland and grassland patches are gradually distributed, the fragmentation is reduced by 76.19%, and the connectivity index is gradually increased. From a spatial perspective, the edge area outside the Fourth Ring Road becomes the most obvious area of fragmentation tendency. Finally, by introducing big data into the landscape layout optimization simulation, a layout optimization simulation method based on big data is constructed and used in the layout optimization simulation of the landscape image layout of the three-dimensional pastoral complex to guide the participatory layout optimization simulation.

## 1. Introduction

Three-dimensional landscape image layout of pastoral complex is an important guarantee for urban ecological environment. By the urban green land, farmland, forestland, and wetland and the waters together form a complex three-dimensional rural landscape image layout that has important ecological service function, in the maintenance of ecological safety and delay species, reducing the heat island effect and maintaining the social and economic sustainable development playing an important role, and has a decisive influence on land ecological environment. A complex three-dimensional rural landscape image layout of ecological benefit, economic benefit, and social benefit also caused the country to attach great importance to

the construction of urban forest and green space construction as the main content; emphasizing the urban forest ecosystem plays an important part in urban development and together with the urban green space system affects the ecological landscape of the city. Farmland, wetland, and water area are also indispensable parts of urban ecosystem, which play an important role in ecological protection and material circulation. Therefore, it is urgent to study the three-dimensional landscape image layout of urban green space, farmland, forestland, and water area and put forward the optimization strategy for the overall urban ecological environment construction and the improvement of ecological benefits.

Theoretical research mainly focuses on the impact of big data on subject teaching, landscape layout optimization

simulation process, and the guidance of data analysis on layout optimization simulation. The actual cases mainly focus on spatial structure analysis, human behavior characteristics, and users' emotional evaluation under the background of big data, mainly using remote sensing, bus ride record, taxi GPS track, Weibo positioning data, network text, and so forth. The research scope involves large-scale scenic spot layout optimization simulation, as well as small- and medium-sized green space system and greenway layout optimization simulation. It can be seen that studies on big data in landscape layout optimization simulation are concentrated in local areas, and no complete system has been proposed for the application of big data technologies and methods in layout optimization simulation program. For example, the concept and characteristics of big data are introduced, and the characteristics of big data are linked with the simulation content of landscape layout optimization [1–3].

Combined with practical cases, the application of location service data in landscape layout optimization simulation in various periods was introduced, and the simulation design method of landscape layout optimization based on location service big data was drawn, providing a good reference and research direction for subsequent researchers [4–6]. In the study of population activity, the quantitative analysis method of thermal diagram was used, which provided a practical reference case for the application of big data in the simulation of landscape layout optimization [7, 8]. Taking Beijing forest park as the research object, the word frequency analysis software is used to collect and analyze the network text, so as to obtain the social service value of different forest parks and introduce the network text into landscape evaluation [9, 10]. By collecting mobile phone location data in scenic area, the behavior rules of tourists are studied to provide a basis for layout optimization simulation management and decision-making in scenic area, and the research on tourists' behavior characteristics based on positioning data is a hot research topic at present [11, 12]. Big data plays a prominent role in the study of urban issues, and it is commonly used to study urban internal space, urban activity space, and urban hierarchy. For example, according to the global Internet map, the world's Internet cities are graded by using domain names and number of users [13, 14]. Use Twine: firewall cracking programs to extract the geographic coordinates and text data, and combined with time series model analyzed in different cities of Twitter posts characteristics and distribution of keywords, found that only a few cities that can offer more time keyword search the information they need and get the city level of network activity and the relationship between 15th and 16th. By analyzing the communication information of 25 million users provided by the Belgian mobile phone operator, the social network was established by using the zip code corresponding to the user's mobile phone bill address, and the communication intensity was studied by using the gravity model, and finally the communication connection and hierarchy system between cities were obtained [15, 16]. By combining mobile network traffic data (directional switching vector and overall network traffic data) with social networking site data (Flickr), the spatial analysis method is adopted to

reveal the activity hotspots of cities and the travel activities and change characteristics of cities and suburbs in different seasons [17, 18]. Through 8 million Flickr locations and image information the central border in London and Chicago metropolitan areas is determined, using GPS and web tools (can record the time and latitude and longitude, speed, distance, and direction) and connecting with Google maps, to simulate the two cities, 76 tourists in Canberra and Sydney, and the action of trajectory, as well as analysis of tourists travel path, change to the mode of transportation, travel, obstacles, and other characteristics [8, 17, 18]. However, there are few studies on landscape development under the background of big data, and the research on big data focuses on the application exploration and decision support of big data and the application of various big data in landscape layout optimization simulation. Big data has brought about new types of technology and equipment to other industries. For landscape architecture, big data has brought about more changes in planning thinking, allowing professionals to rely more on rational analysis of problems. Planning and design is a problem-solving process. Aiming at complex problems, by analyzing a large number of multiple types of data, it proposes a more suitable solution than traditional data analysis. Only in this way can the planning be more objective and rational.

The types of big data involved include cellphone signaling data, satellite positioning data, social network data, and landscape photos. How to mine relevant data under the nonstationary characteristics of big data? Through the investigation of 4950 interviewees on the relationship between the acquisition of high-quality urban green space and the level of leisure and sports activities, it is found that there is no clear relationship between activities and access to green space [19, 20]. Publicly available via the Internet social produce or provide the geographic mark photos, collected from 2100 tourists, 29,443 photographs, with pictures of these geographic staggered build datasets, can help them address these challenges, as well as, for example, destination, traffic layout optimization simulation, and the management development of series of problems such as providing practical use value [21]. Research on the connotation is of three-dimensional rural complex landscape image layout and the relationship between development mode, layout optimization simulation mode and path, industrial layout optimization simulation, and rural revitalization. Based on the analysis of the new urban-rural relationship, the value content, operation mechanism, and guiding strategy of the three-dimensional landscape image layout of pastoral complex are discussed [22, 23]. Layout of a complex three-dimensional rural landscape image, the concept of value, the tertiary industry relations, and operation mode are analyzed in detail; at the same time, summarizing their participation in three-dimensional rural process of landscape image layout project strategy is complex and difficult, and the complex three-dimensional rural landscape image layout in the advantages and problems in the domestic development is demonstrated [24, 25]. The importance of the development of the landscape image layout mode of 3D pastoral complex in

the layout optimization simulation design is listed, and the characteristics of the landscape image layout of 3D pastoral complex are systematically analyzed [26]. The construction contents of the “seven systems” and “seven projects” in the layout of three-dimensional pastoral complex landscape images are introduced in detail, and the layout optimization simulation strategies in terms of industrial system, building rural communities, and joint mechanism of interests are proposed [27, 28].

By learning the data collection and processing technology of big data, the processing mode and guiding method of big data are introduced into the landscape layout optimization simulation, and the processing method of big data is added to the conventional layout optimization simulation program, so as to influence the thinking of layout optimization simulation design and improve the objectivity and scientificness. The main procedure is to introduce the technology of big data collection and processing on the basis of field investigation and site data collection, screen and process the data, and establish a database. By quantifying different types of influencing factors, the thousand-layer cake analysis method is adopted to overlay different data to guide the simulation of landscape layout optimization. The landscape layout optimization simulation method based on the influence of big data was applied to the landscape image layout of 3D pastoral complex, and the similarities and differences were found by comparing the traditional layout optimization simulation method and the layout optimization simulation program of the data-based layout optimization simulation method. The layout optimization simulation design of 3D pastoral complex landscape image layout was studied, focusing on the simulation of road, spatial structure, ecological pattern, and industrial layout optimization of 3D pastoral complex landscape image layout under the premise of ecological environmental protection.

## 2. Big Data Analysis of Three-Dimensional Rural Complex Image Layout Prediction Analysis

The core of landscape pattern optimization of three-dimensional pastoral complex is to realize the optimization of quantity and space. Through quantitative and qualitative research, specific macro and spatial strategies are provided to provide policy and data support for realizing the ideal landscape pattern of three-dimensional pastoral complex. In this study, the system dynamics model and cellular automata model are mainly used as the land simulation and prediction model to carry out the area and space simulation of the landscape pattern optimization of three-dimensional rural complex [29, 30].

### 2.1. Image Layout Predictive Analysis Modeling

- (1) Understand the problem, define the problem, confirm the target, and determine the macro variable factors that affect its change according to the modeling purpose and prediction object. In this

study, the three-dimensional pastoral complex is affected by policy regulation and social economy at the macro level, specifically manifested as economic development, population growth, industrial adjustment, land expansion, and so forth. Therefore, the variable factors of the system dynamics model of three-dimensional pastoral complex should reflect these macro factors, and we call these macro social factors and the system that reflects the area of various three-dimensional pastoral as the complex system of three-dimensional pastoral complex.

- (2) Draw the causal feedback diagram of the system according to the logical relationship between the target and parameters; the causal feedback graph is mainly expressed as arrows and diagrams according to the interaction between variable factors. In addition, positive and negative signs are used to show the interaction between the two factors. That is to say, we need to visualize the relationship between factors in the complex system of three-dimensional pastoral complex in graphical ways such as arrows and signs.
- (3) Describe the relationship between parameters by mathematical formula and establish the system dynamics model. Through debugging in Vensim PLE software, the quantitative relationship between different factors in the 3D pastoral complex system was simulated. By analyzing the interaction mechanism between social economy and the change of the number of covers in three-dimensional pastoral complex, the relationship between them is analyzed mathematically and logically, so as to realize the influence relationship in the whole system.
- (4) Test and verify the model. In order to verify the reasonability of the model, it is necessary to debug the model, compare the relationship between the simulated value and the real value, and usually simulate the data of a certain historical period with the data of the past several stages until the simulation results meet the 10% error range.
- (5) Establish a system mechanics simulation and prediction simulation platform to predict the future area change of the three-dimensional rural complex, design multiple scenarios, and use the model to select and simulate strategies. That is, by changing the parameters of the model and testing the posteffects of various scenarios, the opportunity and approach to improve the behavior of the complex system of three-dimensional pastoral complex are sought. Through the prediction analysis and comprehensive comparison under different scenarios, the relative optimal scheme for the coordinated development of social economy and three-dimensional pastoral complex is sought.

Cellular automata modeling process can be understood as actual three-dimensional rural complex, construction land, and unused land allocation process, refer to previous

high citation rate research, at the same time combined with three-dimensional rural complex in the evolution of the space characteristics, mainly on the basis of land unit is converted to other types of land suitability, neighborhood unit on its transformation, the inheritance of land unit itself and random interference factors to determine, and focus on the transformation between the Mosaic and the three-dimensional rural construction land. The specific modeling process of cellular automata is as follows, as shown in Figure 1.

The statistical results of stakeholders' power dimension of pastoral complex projects are shown in Table 1. It can be seen that the government, social capital, operators, and farmers' cooperatives rank the top four in the power of pastoral complex projects construction, while users, scientific research institutions, financial institutions, and suppliers of materials and equipment rank the bottom four. It is worth noting that financial institutions and users bear more risks in project construction but have less power. This part of project stakeholders is the vulnerable group mentioned in many studies. Farmer cooperatives rank the third in the risk dimension, but their power somewhat decreases. This is because the resources invested by farmer cooperatives may not reach equilibrium with their power, which will lead to the problem of uneven distribution of benefits in the process of project construction and operation. Since the operation stage of the project is relatively important, the operator has more power over the project. The mean and variance are calculated using the formulas in [19].

Pastoral complex is a multifunctional spatial gathering and combination of agricultural production, residence, leisure, and entertainment in a certain regional space and realizes integrated management of spatial layout optimization simulation, construction, and operation. Urban government is the main body of simulation and construction management of spatial layout optimization of pastoral complex. It mainly guides and restricts each participant through policy support and simulation management of spatial layout optimization, as shown in Figure 2.

*2.2. Spatiotemporal Evolution Analysis of Three-Dimensional Pastoral Complex Pattern.* Normalized difference vegetation index (NDVI) is defined as the ratio of the difference between the values of near-infrared band and visible red-light band and the sum of the values of the two bands:  $NDVI = (IR - r) / (IR + R)$ . For TM data, the numbers of near-infrared bands and visible red-light bands correspond to the fourth band (TN4: 0.76–0.9  $\mu m$ ) and the third band (TN3: 0.63–0.69  $\mu m$ ), respectively. NDVI calculation formula is as follows [31]:

$$NDVI = \frac{TN4 - TN3}{TN4 + TN3} \quad (1)$$

As the NDVI value calculated by the above formula is between 0 and 1, in order to match the 8-bit radiation resolution of Landsat, the result is transformed and stretched, so that its gray value is within [0, 255].

Through to the central city in 1992, 2000, 2008, and 2016, four times each type of three-dimensional rural complex remote sensing interpretation results, by Arcmap statistics, get the three-dimensional rural complex interpretation data statistics (Table 2, construction land and unused land) and the three-dimensional rural complex area change contrast figure (Figure 3, construction land and unused land).

It can be seen from Table 2 and Figure 3 that, during the research period from 1992 to 2016, the total area of the three-dimensional pastoral complex in the central city and the area of a single three-dimensional pastoral complex such as grassland, woodland, cultivated land, wetland, and water area all changed to different degrees, and the overall change was relatively obvious. In general, during the study period, the cultivated land and woodland area of the central city always ranked first and second in the area of the three-dimensional pastoral complex, while the wetland, water area, and grassland always ranked second. In 1992, cultivated land was the dominant type of three-dimensional pastoral complex in the central city, but in 2000, the dominant position of cultivated land was slightly reduced. In 2008, woodland became the dominant type of three-dimensional rural complex in the central city, but the advantage was not obvious. In 2016, woodland has become an obvious dominant type of three-dimensional pastoral complex.

Overall, Table 3 shows that, from 1992 to 2016 (study period), the central city's total area of the three-dimensional rural complex had a declining trend; the reason is that forestland, arable land, grassland, and wetland and water appear in different periods of time, and the three-dimensional rural complex type area reduction is higher than the increase, so the total area of the three-dimensional rural complex declined from 1992 to 2016. From the analysis of the area change of different types of three-dimensional pastoral complex, the area of three-dimensional pastoral complex in the central city of Beijing has been undergoing significant dynamic change. The continuous decrease of the overall area of the three-dimensional pastoral complex, the decrease of the large area of cultivated land, the fluctuation of woodland area, and the increase and decrease of the small area of grassland, wetland, and water area have become the typical characteristics of the area evolution of the three-dimensional pastoral complex in the central city since 1992. In the past 24 years, people have changed the way of land use in the central city through urban construction, which will affect the three-dimensional rural complex pattern of the central city. With the expansion of construction land area, the cultivated land area in the central city of Beijing decreases, and the land productivity decreases. With the increase of grassland area and fluctuation of woodland area, the landscape of the study area will become more fragmented, which will not be conducive to the maintenance of biodiversity in the study area and even affect a series of ecological functions such as heat island mitigation and stormwater retention. Therefore, it is necessary to further study the three-dimensional rural complex pattern in the central city of Beijing.

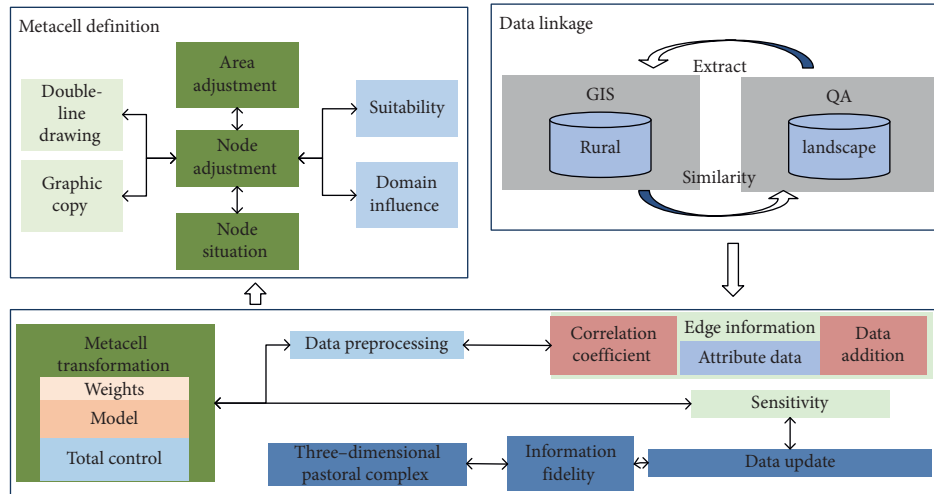


FIGURE 1: Modeling process framework of cellular automata model.

TABLE 1: Data analysis of stakeholder power.

| Project stakeholder                   | Mean  | Variance | Sorting |
|---------------------------------------|-------|----------|---------|
| Government (S1)                       | 4.454 | 1.072    | 1       |
| Farmers' cooperatives (S2)            | 3.878 | 1.232    | 3       |
| Financial institution (S3)            | 2.735 | 1.243    | 10      |
| (S4) by the designer                  | 3.612 | 1.013    | 11      |
| Contractor (SS)                       | 3.836 | 1.205    | 6       |
| The supervisor (S6)                   | 3.856 | 1.401    | 7       |
| Research institutes (S7)              | 2.678 | 1.012    | 4       |
| Operator (S8)                         | 3.977 | 1.075    | 5       |
| Consulting party (integrated)         | 3.229 | 1.124    | 12      |
| Material and equipment supplier (S10) | 3.234 | 1.322    | 9       |
| The user (S11)                        | 2.515 | 1.093    | 8       |
| Social capital (S12)                  | 4.127 | 1.024    | 2       |
| Demolition party (S13)                | 3.304 | 1.315    | 13      |

### 3. Image Layout Optimization Method for Three-Dimensional Rural Landscape Based on Big Data

As an emerging research direction, big data attracts increasing attention. In 2018, there were 25,745 Chinese literatures with big data as the theme, with a sequential growth rate of 21.40%. The study covered ten disciplines, and landscape architecture was not surprisingly affected. Big data in landscape architecture caused a wave of simulations and researches on layout optimization through quantitative information and played a role in providing new technologies or new methods in each stage of landscape layout optimization simulation. In the early analysis of landscape layout optimization simulation, big data mainly introduces new survey methods, expands the data collection methods, and broadens the data collection scope. The additional relevant information provides the foundation for comprehensive, high-precision, and high-granularity research. For example, the data about the characteristics of the users and the emotional information were added to the data collection, and the positioning information was collected through the network. The research of layout optimization simulation

stage focuses on the analysis stage, which is mainly used to discover and mine the relationship between related elements. New data commonly used in layout optimization simulation include location communication data, network text data, and network photos. Positioning because traffic data can reflect the space location and individual behavior, widely used in space function of identification, analysis and use of space structure, road layout optimization simulation of space related problems, such as analysis, as well as the crowd gathers the characteristics of space and time, population distribution, behavior related to human activities such as analysis. Social network is an important channel for public participation layout optimization simulation design due to its wide coverage, fast propagation speed, and good interoperability. Network text data is generated with the development of social networks. By quantifying text and seeking public attention, interest points, and emotional evaluation, it can be used to analyze the advantages and disadvantages of site management and to build an evaluation system. Online photos could not only reflect location data but also reflect the current situation of the site. Superimposing this kind of big visual data on the map can evaluate the space quality, street greening, and road landscape, which is of positive significance to improve the landscape. In the later stage of layout optimization simulation design, big data mainly enriches the feedback platform of the public and establishes an online communication network for the public, layout optimization simulators, and decision-makers through social networks and mobile terminals. Based on the above research, the simulation design framework of landscape layout optimization under the background of big data can be simplified as shown in Figure 4.

First, draw a CAD axis Map, import the CAD Map into the space syntax Depthmap software, use the Convert Drawing Map function to transform the Graph into a layer of Axial Map that can be recognized by the software, and then use the software's Run Graph Analysis function to perform topological calculation on the axis Graph and finally obtain the accessibility Analysis Graph (See Figure 5).



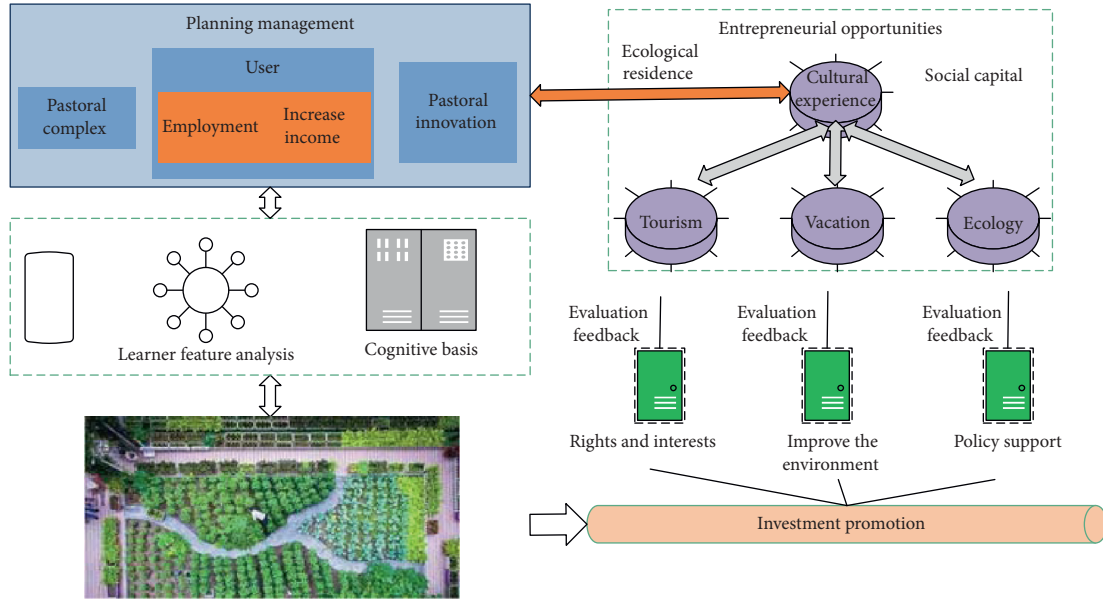


FIGURE 2: Organizational framework and operation mechanism of multiagent pastoral complex.

TABLE 2: Statistical table of land type interpretation data [20].

| In the class        | 1992  |            | 2000  |            | 2008  |            | 2016  |            |
|---------------------|-------|------------|-------|------------|-------|------------|-------|------------|
|                     | Area  | Percentage | Area  | Percentage | Area  | Percentage | Area  | Percentage |
| Grass               | 1838  | 1.68       | 1918  | 1.75       | 2093  | 1.92       | 2247  | 2.07       |
| Arable land         | 32253 | 29.62      | 21172 | 19.45      | 10453 | 9.61       | 6252  | 5.73       |
| Construction land   | 54538 | 50.08      | 70498 | 64.73      | 80493 | 73.88      | 84812 | 77.85      |
| Woodland            | 16418 | 15.06      | 12403 | 11.38      | 14563 | 13.36      | 14175 | 13.02      |
| Wetlands and waters | 3866  | 3.54       | 2928  | 2.68       | 1321  | 1.22       | 1437  | 1.31       |
| Unused              | 18    | 0.02       | 11    | 0.02       | 9     | 0.01       | 6     | 0.01       |

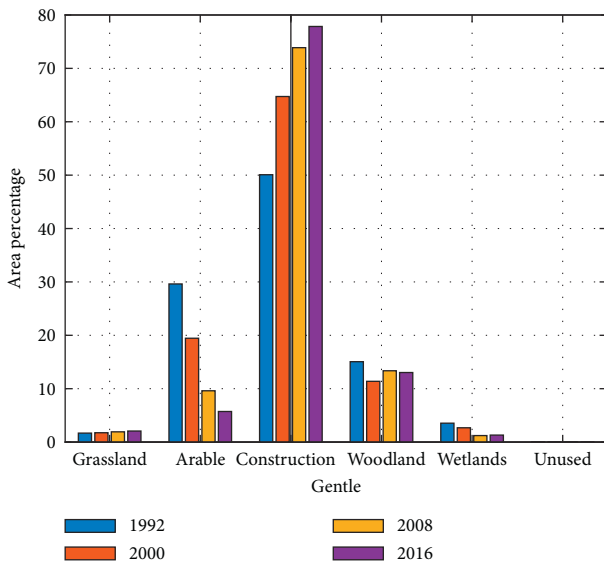


FIGURE 3: Comparison chart of land use area change in the central city from 1992 to 2016.

The area of each land type within the layout optimization simulation range provided by multiple planning is

summarized. According to the ecological capacity of each land type provided by the layout optimization simulation standard of scenic and historic interest areas, the ecological capacity of the land type within the site is defined as shown in Table 4. According to the above formula, the daily space capacity is about 58,100 people, and the annual space capacity is 21,218,500 people.

The area algorithm is adopted to obtain the following formula of tourist mental capacity:

$$Cd = \frac{B}{B_s} \times \frac{R}{r} \quad (2)$$

In the previous equation,  $Cd$  is the psychological capacity of tourism,  $B$  is the total area of sightseeing,  $B_s$  is the per capita area of sightseeing, and  $R$  and  $r$ , respectively, represent the service time and visitor's stay time. The minimum sightseeing area per capita is 50 people. The total area of the site is 9,986,900  $m^2$ , so the daily psychological capacity is 459,300 people, and the annual psychological capacity is 167.68 million people.

It can be seen from Table 5 that, among the three factors, ecological capacity is the smallest, which becomes the determinant to limit the environmental capacity of the site. In layout optimization simulation, the development red line

TABLE 3: Dynamic attitude of land area change of various types.

| In the class                       | 1992–2000      |                  | 2000–2008      |                  | 2008–2016      |                  |
|------------------------------------|----------------|------------------|----------------|------------------|----------------|------------------|
|                                    | Area of change | Dynamic attitude | Area of change | Dynamic attitude | Area of change | Dynamic attitude |
| Grass                              | 81.82          | 0.57             | 175.43         | 1.15             | 154.35         | 0.91             |
| Arable land                        | -11083.4       | -4.4             | -10718.32      | -6.64            | -4201.66       | -5.03            |
| Woodland                           | -4016.62       | -3.07            | 2158.63        | 2.17             | -385.11        | -0.34            |
| Wetlands and waters                | -935.94        | -3.04            | -1607.14       | -6.87            | 116.09         | 1.12             |
| Construction land                  | 15960.29       | 3.64             | 9993.23        | 1.76             | 4319.31        | 0.65             |
| Unused                             | -6.08          | -4.72            | -1.93          | -2.42            | -2.88          | -4.47            |
| Three-dimensional pastoral complex | -15964         | -3.68            | -9993          | -3.26            | -4315          | -1.92            |

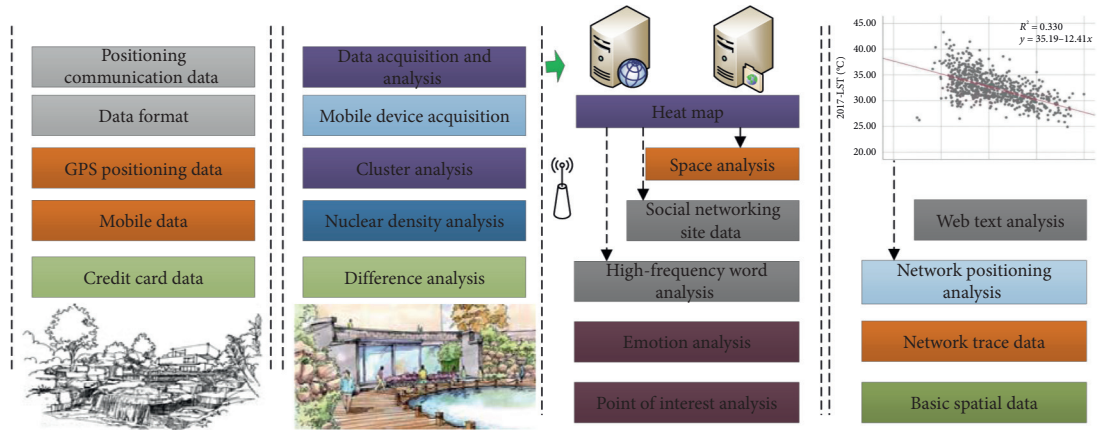


FIGURE 4: Frame diagram of landscape architecture layout optimization simulation design method based on big data.

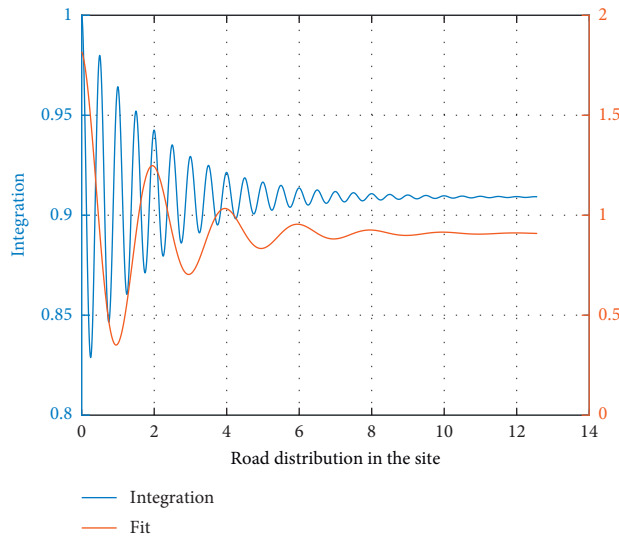


FIGURE 5: Broken line diagram of road axis conformity.

TABLE 4: Summary of the project land area and ecological habitat.

| Land use types   | Area   | Ecological capacity |
|--|--------|---------------------|
| Land for scenic facilities                             | 9.76   | 102                 |
| Rural construction land                                | 124.85 | 103                 |
| Cultivated land (including general and basic farmland) | 366.64 | 505                 |
| Class IV protected woodland                            | 4A6.50 | 1255                |
| River and lake surface                                 | 11.28  | 504                 |

TABLE 5: Environmental capacity of the project.

|                 | Ecological capacity, $C_e$ | Space capacity, $C_s$ | Tourism psychological capacity, $C_p$ |
|-----------------|----------------------------|-----------------------|---------------------------------------|
| Daily capacity  | 5.47                       | 5.82                  | 45.94                                 |
| Annual capacity | 2003.09                    | 2121.86               | 16769                                 |

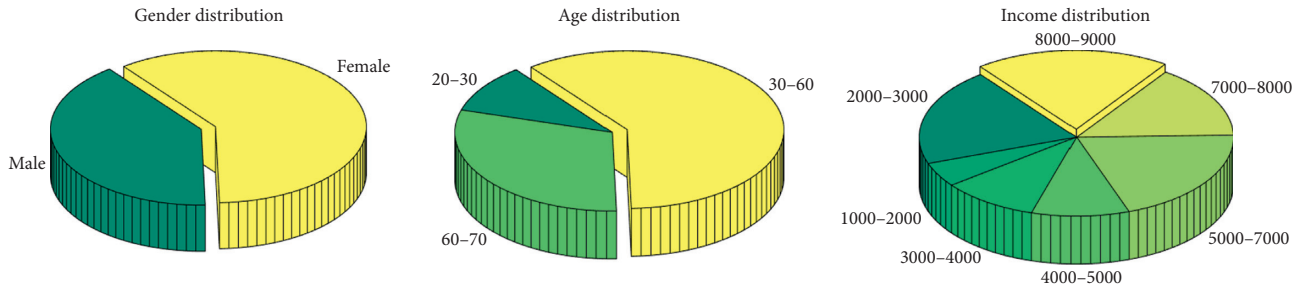


FIGURE 6: Gender distribution, age distribution, and income distribution.

should be the ecological capacity of 54,700 people per day and 20,309,900 people per year.

#### 4. Experimental Verification

Based on the use of big data, this paper conducts a questionnaire to collect basic information of people and their concerns about the pastoral complex. This survey was conducted by WeChat. A total of 65 questionnaires were received, and respondents were distributed in 11 provinces and 1 municipality directly under the central government. The survey content is divided into two parts: one is the basic information of respondents, and the other is the detailed investigation of the construction of pastoral complex. The first part includes basic information such as gender, age, education, occupation, and income and also registers basic information related to tourism, such as the frequency of travel, destination, duration of travel, and fellow travelers. According to the survey results, as shown in Figure 6, the gender distribution of the respondents is more women than men, mainly between 30 and 60 years of age. Most of them have a bachelor's degree and a graduate's degree, and most of them are college students and company employees. Most people go out for an average of two to three times a year, three to four days each time, mainly to scenic spots and tourist cities, and most people go out with friends and family.

The investigation on the content and interest points of pastoral complex is carried out from eight aspects, namely, tourism motivation, types of pastoral complex, elements of pastoral complex, tourism commodities, activities, accommodation forms, simulation points of layout optimization, and regional location. Among the 65 respondents, 64 indicated that they would like to travel to the countryside or take a vacation. Most of them travel to the countryside mainly for rural sightseeing, entertainment, relaxation, and cultural experience. Among the six different types of pastoral complexes, sightseeing and comprehensive and cultural communication pastoral complexes are favored by people. In the nine different elements, natural scenery, local food, farmland landscape, local culture, and living environment are the key points of people's attention. Local specialties, folk customs products, creative

cultural products, and agricultural products account for more than 60% of the tourism commodities. Among the 8 kinds of activities, pastoral tourism, folk custom activities, cultural experience, and farming experience account for up to 50 percent. In the form of accommodation, the proportion of home stay, shared farm, and characteristic hotel is more than 50%. For the elements of pastoral complex, tourism commodities, activities, accommodation forms, and tourism motivations, the proportion of person-time is adopted, as shown in Figure 7.

As for the project positioning of building a healthy and nourishing rural complex with urban backyard garden, all respondents expressed that it was reasonable, as shown in Figure 8. The project positioning of the survey rationality evaluation highlights the characteristics of the site, mainly because the project positioning combines local resources; in response to national policies, development under environmental protection conditions can also drive local development: implementation feasibility is higher.

According to the classification of S and R, public participation can be divided into three types and eight levels, from false (nonfalse) participation to substantive participation, and the degree of participation gradually increases (see Table 6). At present, the level of conventional participation is mainly informed participation, and the level of participation is relatively shallow. In this paper, a participatory layout optimization simulation design program is established by collecting big data and investigation. In the preliminary analysis, the opinions of villagers and netizens were investigated. In the layout optimization simulation fully consider the opinions; after the layout optimization simulation is completed, people's ideas will be understood in the form of video and questionnaire. The level of public participation includes informed participation, cooperative participation, and decision-making participation, which greatly improves the level of participation.

The participatory layout optimization simulation method is introduced in the layout optimization simulation of rural complex, which is helpful to build the communication platform. Everyone has different choices due to

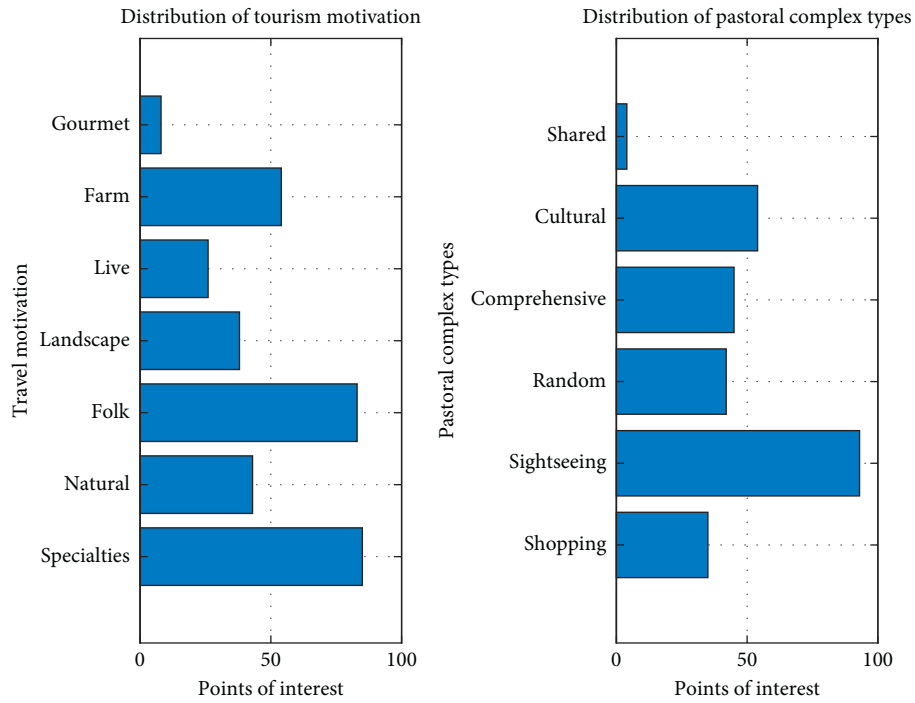


FIGURE 7: Distribution of tourism motivation and distribution of pastoral complex types.

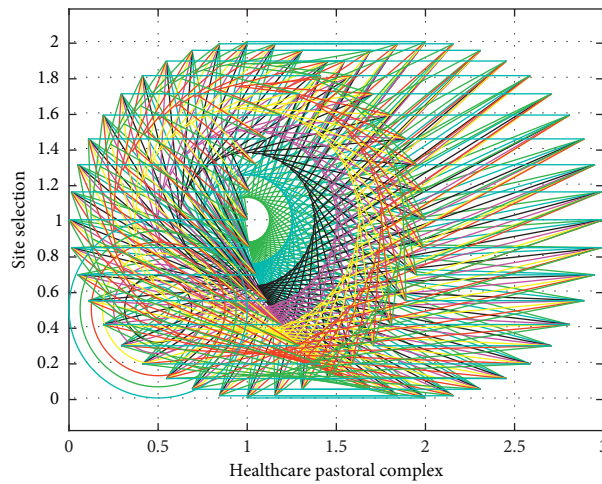


FIGURE 8: Proportion chart of rationality evaluation of project positioning.

TABLE 6: Public participation levels of SR.

|                               |                           |                              |
|-------------------------------|---------------------------|------------------------------|
| Level of public participation |                           |                              |
|                               | False participation       | Operational participation    |
|                               |                           | Educational participation    |
| Public participation          | Symbol in                 | Informed participation       |
|                               |                           | Consultative participation   |
|                               | Substantial participation | Restricted participation     |
|                               |                           | Cooperative participation    |
|                               |                           | Representative participation |

different living environment, education level, working environment, and so forth. Participatory layout optimization simulation can provide a platform for communication,

which can be used to express ideas and guide layout optimization simulation. It is also a platform for mutual learning. Professionals have professional knowledge reserves

but lack the understanding of site information. Although the public have no professional knowledge, they have a good grasp of the site information, especially the cultural aspects that cannot be consulted on the Internet or in the literature. The communication between the two is conducive to a comprehensive understanding of the site, as well as layout optimization simulation design with cultural characteristics of the scheme.

## 5. Conclusion

The landscape layout optimization simulation design based on the background of big data can apply big data and big data technology to every stage of layout optimization simulation. In the preliminary analysis, big data provides layout optimization simulation with a wider range of more comprehensive basic data and information about crowd concerns. In addition to the new data, it also introduces a lot of analysis methods, such as regression analysis, factor analysis, emotion analysis, time series analysis, and nuclear density analysis, providing a channel for in-depth mining of the information behind the data. In the late stage of layout optimization simulation, the visual presentation mode enables more people to understand the intention and effect of layout optimization simulation. In combination with the public feedback platform of big data, modification suggestions are provided for layout optimization simulators. In the process of constructing layout optimization simulation framework, the authors found that although big data brought about changes in layout optimization simulation thinking, update of analysis methods, new technologies, and other influences, big data mainly played the role of supporting analysis in the whole layout optimization simulation. We should embrace the impact of big data with an open mind and at the same time be rational in its use. Rationally view the positioning of big data in layout optimization simulation; Face up to the difference between rational and perceptual analysis methods and combine them; only in this way can the power of big data be brought into play. Against the background of big data, a simulation program for layout optimization of pastoral complex is constructed, in which the application of big data in pastoral complex mainly focuses on the early analysis and the late feedback. In the early analysis, the global consciousness of big data plays a huge role in the analysis of the overall layout, which enables the layout optimization simulator to consider the site in a larger scope. At the same time, based on big data analysis, it can understand and master the interest points and basic features of the public, which is conducive to providing targeted activities. On the basis of previous analysis, the influence of big data on layout optimization simulation is mainly focused on spatial analysis and layout optimization simulation, such as function partition, spatial structure layout optimization simulation, and spatial quantization. Big data plays a guiding role. The greater influence of big data on layout optimization simulation of rural complex lies in that it provides a communication platform for layout optimization simulation designers, farmers, and decision-makers, deepening the degree of participatory layout optimization simulation.

Affected by professional optimization of the 3D pastoral landscape, research time and way of thinking, the depth and breadth of model construction, research methods, and problem-solving ideas are obviously insufficient, and the angle of innovation is not enough, which may lead to certain problems in research bias.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This work was supported by 2019 National Office of Philosophy and Social Sciences Ethnology project: the evolution of the Miao Ethnic Costume Culture and the Transformation of Contemporary Innovation in Wuxi Valley (Project no. 19BMZ083).

## References

- [1] Y. Yang, D. Z. Luk, H. Zhou, C. Yan, D. Zhou, and X. Zeng, "Layout decomposition Co-optimization for hybrid E-beam and multiple patterning lithography," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 35, no. 9, pp. 1532–1545, 2016.
- [2] M. R. Amer, S. Yousefi, and S. Todorovic, "Monocular extraction of 2.1D sketch using constrained convex optimization," *International Journal of Computer Vision*, vol. 112, no. 1, pp. 23–42, 2015.
- [3] Q. Yu, Y. D. Yue, M. FangKai, H. Ma, Q. Zhang, and Y. Huang, "Optimization of ecological node layout and stability analysis of ecological network in desert oasis: a typical case study of ecological fragile zone located at Deng Kou county(inner Mongolia)," *Ecological Indicators*, vol. 84, no. 1, pp. 304–318, 2018.
- [4] S. Yamauchi, S. Yamasaki, and K. Yaji, "Multidisciplinary layout design optimization method using multi-fidelity analysis model inspired by the explicit method (application to layout design problem considering a heat dissipation characteristic)," *Transactions of the Jsme*, vol. 83, no. 855, pp. 320–337, 2017.
- [5] M. Galun, I. R. Basri, and I. Yavneh, "Review of methods inspired by algebraic-multigrid for data and image analysis applications," *Numerical Mathematics: Theory, Methods and Applications*, vol. 8, no. 2, pp. 283–312, 2015.
- [6] P. Buccella, H. C. Stefanucci, R. IskanderMoursy, J.-M. Sallese, and M. Kayal, "Methodology for 3-D substrate network extraction for SPICE simulation of parasitic currents in smart power ICs," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 35, no. 9, pp. 1489–1502, 2016.
- [7] S. Koranne, "Design and analysis of silicon photonics wave guides using symbolic methods," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 34, no. 3, pp. 341–353, 2015.

- [8] S. M. Wang, S. Ma, and W. Y. Duan, "Seakeeping optimization of trimaran outrigger layout based on NSGA-II," *Applied Ocean Research*, vol. 78, pp. 110–122, 2018.
- [9] T. Liu, J. H. Zhu, and W. H. Zhang, "Integrated layout and topology optimization design of multi-component systems under harmonic base acceleration excitations," *Structural and Multidisciplinary Optimization*, vol. 59, no. 5, pp. 1053–1073, 2019.
- [10] R. Chen, V. K. Singapura, and V. K. Prasanna, "Optimal dynamic data layouts for 2D FFT on 3D memory integrated FPGA," *The Journal of Supercomputing*, vol. 73, no. 2, pp. 652–663, 2017.
- [11] J. Lim and K. M. Lee, "Design of electromagnetic actuators using layout optimization with distributed current source models," *IEEE/ASME Transactions on Mechatronics*, vol. 20, no. 6, pp. 1–10, 2015.
- [12] P. Buccella, C. Stefanucci, and J. M. Sallese, "Simulation, analysis, and verification of substrate currents for layout optimization of smart power ICs," *IEEE Transactions on Power Electronics*, vol. 31, no. 9, pp. 53–67, 2016.
- [13] A. Chvála, J. Marek, A. Šatka et al., "Analysis of multifinger power HEMTs supported by effective 3-D device electro-thermal simulation," *Microelectronics Reliability*, vol. 78, pp. 148–155, 2017.
- [14] C. Sun, W. J. Zhao, and W. Huang, "New insight into complex mode matching method for modeling and simulation of surface-emitting grating couplers," *Journal of Lightwave Technology*, vol. 37, no. 3, pp. 839–844, 2019.
- [15] C. J. Smith, M. Gilbert, and F. Derguti, "Application of layout optimization to the design of additively manufactured metallic components," *Structural and Multidisciplinary Optimization*, vol. 54, no. 5, pp. 1297–1313, 2016.
- [16] L. Meng, J. Wu, and J. Dong, "Spatial differentiation and layout optimization of rural settlements in hill ecological protection area," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 33, no. 10, pp. 278–286, 2017.
- [17] L. He and M. Gilbert, "Rationalization of trusses generated via layout optimization," *Structural and Multidisciplinary Optimization*, vol. 52, no. 4, pp. 677–694, 2015.
- [18] S. Peng and J. An, "Prediction and verification of risk loss cost for improved natural gas network layout optimization," *Energy*, vol. 148, no. 4, pp. 1181–1190, 2018.
- [19] M. I. Alam and R. S. Pant, "Multi-objective multidisciplinary design analyses and optimization of high altitude airships," *Aerospace Science and Technology*, vol. 78, no. 6, pp. 248–259, 2018.
- [20] J. Y. Han, Y. M. Kim, Y. L. Rhie, S. H. Ahn, H. Choi, and M. H. Yun, "Design optimization of control layout for naval mfc (multi-function console) using A modified layout analysis method," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 59, no. 1, pp. 1351–1355, 2015.
- [21] O. L. Sgrott, D. Noriler, and H. F. Meier, "Cyclone optimization by COMPLEX method and CFD simulation," *Powder Technology*, vol. 277, pp. 11–21, 2015.
- [22] X. Shang, J. Zhou, and F. Zhuo, "Analysis of crack for complex structural parts and simulation optimization during hot forming," *International Journal of Advanced Manufacturing Technology*, vol. 80, no. 1–4, pp. 373–382, 2015.
- [23] Y. Chen, H. W. Wang, and S. H. Zhang, "Forging process design and simulation optimization of a complex-shaped aluminium alloy component," *Materials Science Forum*, vol. 941, pp. 784–789, 2018.
- [24] G. Liuzzi, V. S. Lucidi, and V. Piccialli, "Global optimization of simulation based complex systems," *Uncertainty Management in Simulation-Optimization of Complex Systems*, vol. 59, pp. 173–202, 2015.
- [25] A. Rheinlaender, U. Leser, and G. Graefe, "Optimization of complex dataflows with user-defined functions," *ACM Computing Surveys*, vol. 50, no. 3, pp. 38.1–38.39, 2017.
- [26] H. Kulhari, M. K. D. Pooja, and A. S. Chauhan, "Optimization of carboxylate-terminated poly (amidoamine) dendrimer-mediated cisplatin formulation," *Drug Development and Industrial Pharmacy*, vol. 41, no. 2, pp. 232–238, 2015.
- [27] R. J. Park and S. H. Byun, "Optimization of a neutron long counter design by monte carlo simulation," *Health Physics*, vol. 117, no. 3, pp. 300–305, 2019.
- [28] H. Z. Zhang, X. Liu, and W. J. Yi, "Reinforcement layout optimization of members under complex stress states and multiple loading cases," *Journal of Hunan University (Natural Ences)*, vol. 41, no. 9, pp. 42–47, 2014.
- [29] R. Brogna, J. Feng, and J. N. Srensen, "A new wake model and comparison of eight algorithms for layout optimization of wind farms in complex terrain," *Applied Energy*, vol. 259, pp. 114189–114202, 2019.
- [30] Y. D. Sun, Q. R. Chen, and W. J. Sun, "Numerical simulation of extrusion process and die structure optimization for a complex magnesium doorframe," *The International Journal of Advanced Manufacturing Technology*, vol. 80, no. 1–4, pp. 495–506, 2015.
- [31] M. Freitag and T. Hildebrandt, "Automatic design of scheduling rules for complex manufacturing systems by multi-objective simulation-based optimization," *CIRP Annals*, vol. 65, no. 1, pp. 433–436, 2016.

## Research Article

# Improved Newton Iterative Algorithm for Fractal Art Graphic Design

Huijuan Chen and Xintao Zheng 

*Academy of Arts, Hebei GEO University, Shijiazhuang 050031, Hebei, China*

Correspondence should be addressed to Xintao Zheng; zhanghua@hgu.edu.cn

Received 19 October 2020; Revised 13 November 2020; Accepted 18 November 2020; Published 29 November 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Huijuan Chen and Xintao Zheng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Fractal art graphics are the product of the fusion of mathematics and art, relying on the computing power of a computer to iteratively calculate mathematical formulas and present the results in a graphical rendering. The selection of the initial value of the first iteration has a greater impact on the final calculation result. If the initial value of the iteration is not selected properly, the iteration will not converge or will converge to the wrong result, which will affect the accuracy of the fractal art graphic design. Aiming at this problem, this paper proposes an improved optimization method for selecting the initial value of the Gauss-Newton iteration method. Through the area division method of the system composed of the sensor array, the effective initial value of iterative calculation is selected in the corresponding area for subsequent iterative calculation. Using the special skeleton structure of Newton's iterative graphics, such as infinitely finely inlaid chain-like, scattered-point-like composition, combined with the use of graphic secondary design methods, we conduct fractal art graphics design research with special texture effects. On this basis, the Newton iterative graphics are processed by dithering and MATLAB-based mathematical morphology to obtain graphics and then processed with the help of weaving CAD to directly form fractal art graphics with special texture effects. Design experiments with the help of electronic Jacquard machines proved that it is feasible to transform special texture effects based on Newton's iterative graphic design into Jacquard fractal art graphics.

## 1. Introduction

Fractal geometry is often used to describe irregular things in nature. The well-known Euclidean geometry describes objects composed of points, straight lines, common polygons and curves in two dimensions, and boxes and surfaces in three dimensions [1, 2]. Most common man-made objects can be described by Euclidean geometry, such as books, desks, lighthouses, houses, and other buildings. However, many natural objects in nature cannot be described by conventional geometric figures, such as clouds and coastlines [3]. The emergence of fractal geometry provides a new perspective for describing natural objects. In fact, it is difficult to judge the difference between two clouds and two coastlines only from the structure and shape, because these natural objects have self-similarity. Fractal geometry uses the idea of fractal dimensions to describe the difference in this

characteristic and maps objects with different dimensions [4].

In foreign countries, the research on fractal art is mainly based on scientists, and artists' attention to this field is not very common [5]. The research of fractal art has always had its own system and development history, which is particularly prominent in design education [6]. Reason education has been mentioned in art education for a long time, and geometry is even regarded as one of the foundations of art design teaching [7]. Joseph Albers, Max Bill, and others have cited many geometric principles in graphic design to guide operation and design [8, 9]. A large part of their courses are graphic creation based on mathematical prototypes. In China, the research on fractal art mainly focuses on artists, and scientists pay little attention [10]. This is a manifestation of the different understandings of fractal art research at home and abroad. Relevant scholars explained how Julia set,

Mandelbrot set, and Newton fractal set adjust the number and shape of petals in clothing design and extracted texture information of different flower types, focusing on analyzing the relationship between Julia set flower types and various parameters [11]. Researchers proposed that works of art should represent nature and work like nature [12]. Science is also trying to explain the laws that determine nature. Technology provides appropriate tools for both parties to achieve common goals [13]. Fractal art is at the core of the triangular relationship between art, science, and technology. Fractal geometry and chaos theory bring new perspectives to art. The study of fractal features provides broad possibilities for art development [14]. Relevant scholars have used artificial neural networks to realize the fusion of natural scene image content and classic painting style and transfer the classic painting style to the content image [15]. Through parameter adjustment, the generated image meets people's aesthetic standards. As an important branch of deep learning, style transfer has been favored by many scholars [16]. In terms of practical application, because theory and practice have not been well integrated and due to lack of understanding of fractals, there are very few researches on the rules of graph generation and the application of the rules [17]. Only the clothing and textile industry is involved and requires fractal technology. Higher industries such as movies and 3D games have no real substantive applications for the time being. In addition, due to the disconnection between science education and art education in China, the domestic understanding of fractal graphics is still at the initial stage. Scientists are not paying enough attention to this kind of artistic performance, so it has not attracted more attention. The research on fractal art lacks knowledge not only in theory but also in material support. Therefore, the road to popularization of fractal theory is still very difficult [18–20]. How to further promote fractal theory to make people realize its material value and economic benefits, or to use fractal art image as a new art form, requires our further efforts.

Newton's iterative graph is generated from nonlinear dynamics theory by changing its mathematical model and related parameters. In the calculation process, the Gauss-Newton iteration method is used to optimize the selection of the initial value of the Gauss-Newton iteration method. The optimization method makes full use of the characteristics of the actual problem of target positioning and combines the advantages of the Gauss-Newton method to perform local fast search. The situation where the Gauss-Newton method does not converge or only converges to the local optimal solution is avoided, and a more ideal positioning result is obtained. Four Newton iterative transformation forms are proposed to produce different shapes of graphics, which can affect the texture effect of the fractal art graphics surface. Mathematical morphology processing of Newton's iterative graphics with the help of MATLAB makes the pixel points of the graphics directly correspond to the tissue points, making them directly into fractal art graphics and presenting a variety of special texture effects, and we design experiments with the help of electronic Jacquard machines. In order to

design special texture effects of fractal art graphics, a new way for reference is explored.

The rest of this article is organized as follows. Section 2 analyzes the related theories of fractal art graphics. Section 3 constructs an improved Newton iteration algorithm. Experiments and discussions were conducted in Section 4. Section 5 summarizes the full text.

## 2. Theories Related to Fractal Art Graphics

*2.1. Characteristic Analysis of Fractal.* A fractal is a collection of some "complex" points in some simple spaces. This collection has some special properties. First, it is a compact subset of the space where it is located and has the typical geometric characteristics listed below:

- (i) The fractal set has proportion details at any small scale, or it has a fine structure.
- (ii) The fractal set cannot be described by traditional geometric language. It is neither the trajectory of points that satisfy certain conditions nor the solution set of some simple equations.
- (iii) The fractal set has a certain self-similar form, which may be approximate self-similar or statistical self-similar.
- (iv) The "fractal dimension" (defined in some way) of a fractal set is generally greater than its topological dimension.
- (v) In most interesting situations, the fractal set is defined by a very simple method and may be generated by iterations of transformations.

For a variety of different fractals, some may have all the above properties at the same time, some may only have most of them, and some have exceptions to certain properties, but this does not affect us calling this set a fractal. It should be pointed out that most of the fractals involved in nature and various applied sciences are approximate. When the scale is reduced to the size of the molecule, the fractality disappears, and strict fractal exists only in theoretical research.

Fractals are generally divided into two categories, deterministic fractals and random fractals. If multiple iterations of the algorithm still produce the same fractal, this fractal is called a deterministic fractal. Deterministic fractals are repeatable. Even though some randomness may be introduced in the generation process, the final graph is deterministic. Random fractal refers to the fact that although the rules for generating fractals are determined, they are affected by random factors. Although the fractals generated by each generation process can have the same complexity, the shape will be different. Although random fractals also have a set of rules, the introduction of randomness during the generation process will make the final graph unpredictable. That is, the graphics generated by the two operations at different times can have the same fractal dimension, but the shape may be different, and random fractals are not repeatable. The frame diagram of the fractal graphic design program is shown in Figure 1.



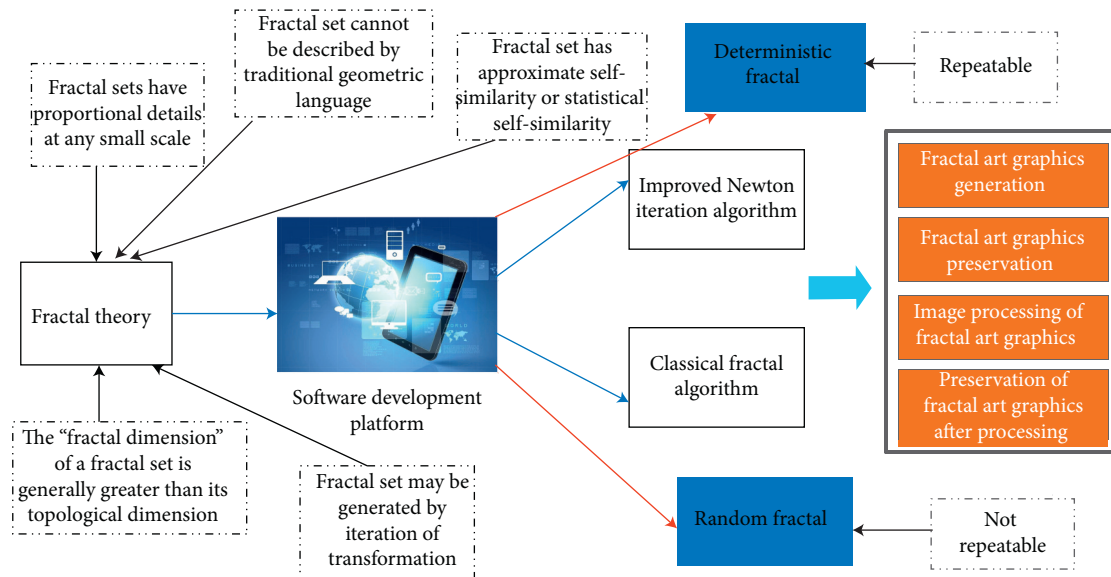


FIGURE 1: Framework diagram of fractal graphic design program.

**2.2. The Difference between Fractal Geometry and Euclidean Geometry.** To explain the difference between fractal geometry and Euclidean geometry, first we introduce the characteristics of Euclidean geometry. Euclidean geometry is a study of regular geometric figures. The so-called regular geometric figures are familiar points, straight lines, and line segments; squares, rectangles, trapezoids, rhombuses, various triangles and regular polygons on planes, and planes; and cubes, cuboids, and regular tetrahedrons in space. The other type is geometric figures composed of curves or surfaces, circles and ellipses on a plane, spheres, ellipsoids, cylinders, cones, and truncated cones in space. The dimensions (Euclidean dimension) of these points, lines, screen graphics, and space graphics are 0, 1, 2, and 3, respectively. The geometric measurement of regular geometric figures refers to the measurement of length, area, and volume.

The graphics studied by fractal geometry are more complex or more realistic than those studied by Euclidean geometry. Its important feature is that it has no characteristic length, and the lines or surfaces that make up its shape are not smooth and nondifferentiable. For example, clouds are not spherical, mountains are not conical, coastlines are not arcs, tree bark is not smooth, and even lightning does not traverse the sky in a straight line. These irregular geometric shapes are difficult to describe with straight lines, smooth curves, and smooth curved surfaces in Euclidean geometry. Therefore, the research object of fractal geometry is a kind of irregular geometric shapes with no characteristic length. Although this kind of object cannot be processed by classical Euclidean geometry, it has “good” properties. To facilitate research, important idealized assumptions are often made; that is, it is assumed to be self-similar. Self-similarity means that if a part of the figure to be considered is enlarged, its shape is the same as the whole. Although these assumptions are too simplistic, only then can we study them while still being suitable for the purpose of the

application. Of course, no real structure will remain the same after an infinite number of repeated amplifications. In principle, self-similarity is only approximately reflected in the application.

There is no strict definition of the characteristic length. Generally, the length that can represent the geometric characteristics of an object is called the characteristic length of the object, such as the radius of a sphere, the side length of a cube, and the height of a person; these are the characteristic lengths of various objects, and they well reflect the geometric characteristics of these objects. For the shapes of objects with characteristic lengths, even if they are slightly simplified, as long as their characteristic lengths remain unchanged, their geometric properties will not change much. In other words, for this type of object, you can use geometrically well-known simple shapes such as rectangles, cylinders, and spheres to combine them, and they can closely resemble their structures. For objects that do not have a characteristic length, the characteristic is that they cannot or are difficult to measure with conventional geometric scales.

**2.3. Fractal and Chaos.** Fractals often show irregular representations, but this does not mean that they are absolutely irregular. Fractals have the characteristics of “self-similar”; that is, they take any part of the fractal figure and enlarge it appropriately, and you can still get a similar image to the original whole figure.

The object described by chaos has an infinite self-similar structure and also has an irregular representation but actually has an infinite self-similar nested structure. In this way, the research on “fractal” and “chaos” has moved towards convergence. We can see the fact that there is a chapter on “fractal” in the book titled “Chaos,” but in the book titled “Fractal,” there is another chapter on “Chaos.” “Fractal” and “Chaos,” the two theories developed from different angles, converge on “self-similarity.”

Nonlinear scientific research turns people's understanding of "normal" things and "normal" phenomena to the exploration of "abnormal" things and "abnormal" phenomena. "Multimedia" technology is a new "unconventional" method used to encounter a large number of "unconventional" phenomena in the process of information storage, compression, conversion, and control. Chaos breaks the various "singular attractors" phenomenon that the deterministic equation determines the motion of the system by the initial conditions.

Chaos comes from a nonlinear dynamic system, and the dynamic system describes an arbitrary process that develops and changes over time. Such systems arise from all aspects of life. The research purpose of the dynamic system is to predict the final development result of the "process." However, even the simplest dynamic system with only one variable will have an essentially random characteristic that is difficult to predict. The sequence produced by successive iterations of a point or a number in a dynamic system is called an orbit. If a small change in the initial conditions causes the corresponding orbit to change only slightly within a certain number of iterations, the dynamic system is stable. At this time, the orbit arbitrarily close to the given initial value may be far from the original orbit. Therefore, it is extremely important to understand the set of unstable points in a given dynamic system. The set of all points whose orbits are unstable is the chaotic set of this dynamic system, and small changes in the parameters of the dynamic system can cause rapid changes in the structure of the chaotic set. This kind of research is extremely complicated, but, with the introduction of a computer, you can visually see the structure of this chaotic set and see whether it is a simple set or a complex set and how it changes as the dynamic system itself changes. It is from here that fractal enters the chaotic dynamic system research.

Chaos mainly discusses the unstable divergence process of a nonlinear dynamic system, but the system always converges to a certain attractor in the phase space, which is very similar to the generation process of fractals. Chaos mainly discusses the behavioral characteristics of the research process, while fractal pays more attention to the study of the structure of the attractor itself. At the same time, chaos and fractal rely heavily on the advancement of computers, which poses a challenge to the traditional concept of pure mathematics. It also greatly stimulated the interest and understanding of scientists and the public and played a role in promotion. The consistency of fractal and chaos is not accidental. In the computer image of chaos set, it is often the set of points with unstable orbit that forms the fractal. So these fractals are given by an exact rule. They are a chaotic set of dynamical systems and various strange attractors. Therefore, the beauty of fractal images is the beauty of chaotic collections, and the study of fractal images is part of the study of chaotic dynamics.

*2.4. Method of Generating Fractal Graphics.* The  $L$  system is a set of methods to describe plants and trees proposed from the perspective of plant morphology. At the beginning, it only focused on the topological structure of plants, that is,

the neighboring relationship between plant components. After years of research, geometric explanations were added to the description process. The high simplicity and multilevel structure of this system provide an effective theory and method for describing the morphological and structural characteristics of the growth and reproduction process of plants and trees. Not only can the  $L$  system describe plants but also its composition method can be used to draw all kinds of regular fractal curves and other shapes. Figure 2 shows the calculation framework of the fractal dimension value generated by fractal graphics.

Iteration Function System (IFS) is an important branch of fractal geometry, and it is also one of the most vital and promising fields in fractal images. IFS is a fractal configuration system. Aiming at this system, a set of theories was proposed, a series of algorithm rules were developed, and they were used in many aspects. The theory of IFS includes compression mapping, metric space, existence of invariant compact sets, and measurement theory. The iterative function system has great advantages in the modeling of a large class of objects, especially the advantages of computer simulation of natural scenery. Because of this, IFS has a wide range of applications in graphics. Among them, the research of visualization technology has been extended from 2D fractal to 3D fractal objects; the self-similar fractal image researched by IFS expands its application range, and the IFS transformation need not be limited to affine transformation. For the geometric transformation of the original graphics, the linear transformation in IFS is extended to the nonlinear transformation; for the discussion of the computer generation of natural scenery, the modeling method is also extended from two-dimensional to three-dimensional.

The fractal set of complex dynamical system mainly includes Mandelbrot set and Julia set. Mandelbrot set is the most famous fractal set in fractals. Julia sets are an iteration of polynomials and rational functions. Both the Mandelbrot set and the Julia set are sequences of points obtained by repeated iterations in the complex plane.

The Mandelbrot set is a general outline of the Julia set, and the Julia set is the boundary of the Mandelbrot set. The beautiful images presented in front of people by Mandelbrot and Julia impressed the artists. The benefits have broad application prospects. People's research on Julia set and its extension includes generation algorithm, related demonstration, three-dimensional fractal graph generation, and its further extension; the research on the mapping of Julia set also has further development, including high-order Julia set generation. The study of methods extends the quadratic complex mapping to higher-order complex mapping; the second-order Julia set algorithm-escape time algorithm, random inverse function algorithm, and rotating escape time algorithm are extended to higher-order and generalized Julia sets and the fractal image of Julia set. Through computer experiment method, the research of Julia set has been extended to transcendental function; this research field of artistic design based on fractal and Julia set image has set up another pass for people.

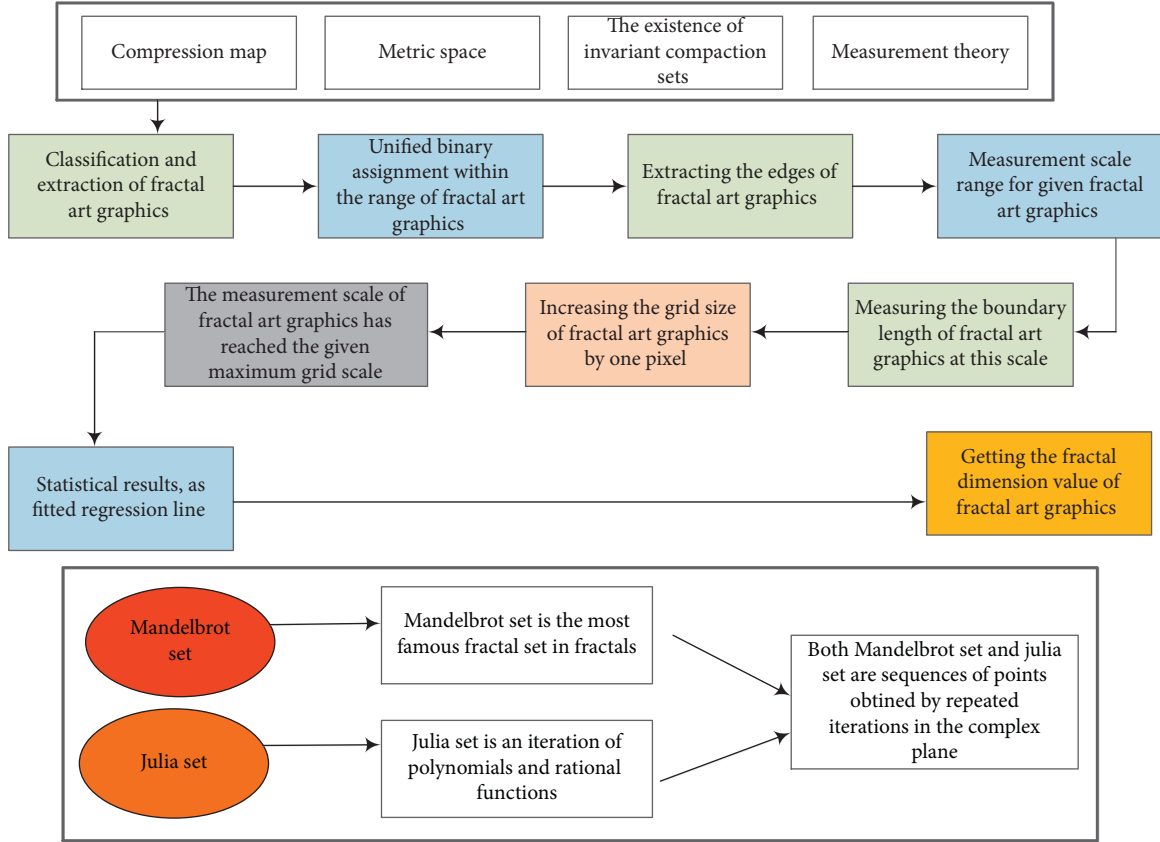


FIGURE 2: The calculation framework of fractal dimension values generated by fractal graphics.

### 3. Improved Newton Iteration Algorithm

**3.1. Gauss-Newton Iteration Method.** Gauss-Newton iteration method is used to solve nonlinear regression problems. After setting the initial value, through multiple iterations, the regression coefficients are modified to obtain the optimal solution of the equations. The basic idea of the Gauss-Newton iterative method is to use Taylor series expansion to approximately replace the nonlinear regression model, and then, after several iterations, the regression coefficients are constantly revised to make the regression coefficients constantly approach the best regression coefficients of the nonlinear regression model. The goal is to minimize the residual sum of squares of the original model. The Gauss-Newton algorithm is an algorithm for solving nonlinear least squares problems. It can be seen as Newton's method to find a minimum function variant. It is used to minimize the sum of the square values of the function. In nonlinear regression, the parameters in the model are sought after, making the model consistent with existing observations, such as nonlinear least squares problems.

Given  $m$  functions  $R = (R_1, \dots, R_m)$  of  $n$  variables  $\alpha = (\alpha_1, \dots, \alpha_n)$ , where  $m \geq n$ , the Gauss-Newton iterative algorithm finds the least square sum:

$$s(\alpha) = \sum_{i=1}^m r_i^2(\alpha). \quad (1)$$

We iterate from the set initial value

$$\alpha^{(s+1)} = r\alpha^{(s)} J_r^T (J_r J_r^T)^{-1} + \alpha^{(s)}. \quad (2)$$

Among them, the Jacobian matrix is

$$J_r = \frac{\partial r_i(\alpha^{(s)})}{\partial \alpha_j}. \quad (3)$$

We set the system of equations to

$$f_i(x, y) = [(x - x_i)^2 + (y - y_i)^2]^{1/2} - [(x - x_{i+1})^2 + (y - y_{i+1})^2]^{1/2} + vt_{i,i+1}. \quad (4)$$

Its Jacobian matrix is

$$G(x, y) = \begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \\ \frac{\partial f_3}{\partial x} & \frac{\partial f_3}{\partial y} \end{pmatrix} = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \\ C_{31} & C_{32} \end{pmatrix}. \quad (5)$$

When the Jacobian matrix is a nonsingular matrix (the determinant is not zero), the target coordinate iteration formula is

$$\begin{pmatrix} x^{(k+1)} \\ y^{(k+1)} \end{pmatrix} = \begin{pmatrix} f_1(x^{(k)}, y^{(k)}) \\ f_2(x^{(k)}, y^{(k)}) \\ f_3(x^{(k)}, y^{(k)}) \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \\ \frac{\partial f_3}{\partial x} & \frac{\partial f_3}{\partial y} \end{pmatrix} + \begin{pmatrix} x^{(k)} \\ y^{(k)} \end{pmatrix}. \quad (6)$$

**3.2. Selection of Initial Value.** The Gauss-Newton iteration method is greatly affected by the initial value. The three-node set in the target positioning system used in this article can form a triangular array, and the function value of the Jacobian matrix in the corresponding direction is calculated by calculating the points on the outer extension line of each side of the triangle. The function value no longer changes.

At points on the straight line, the value of the Jacobian matrix is 0. At this time, the Jacobian matrix function takes the extreme value. The Jacobian matrix function takes the minimum value on the left side of the point and the maximum value on the right side, but the value of the Jacobian matrix function remains unchanged. Therefore, when the Gauss-Newton iteration method is used to solve the problem and when the initial value and the real calculation result are in the same range (referring to the same side of the extreme value point), the optimal solution can be obtained after several iterations. When the initial value and the true value are not in the same range, more accurate calculation results cannot be obtained, and only iterative convergence can lead to a local optimal solution.

Based on the idea of genetic algorithm, the selection of the initial value of Gauss-Newton iteration method is optimized. Genetic algorithm is a nondeterministic quasi-natural algorithm. Genetic algorithm is a random algorithm that uses natural selection and genetic mechanism in nature. The main idea of the algorithm is to simulate heredity, mutation, and crossover in natural selection. We select ideal individuals and recombine them through genetic operators to generate a new set of candidate solution groups, until the optimal solution or better solution that meets the setting is obtained. Genetic algorithm provides a new method for the selection of the initial value of the Gauss-Newton iteration method. In the calculation of the Gauss-Newton iteration method, a group of candidate points are selected to participate in the iteration, and the reasonable selection parameters are selected according to the fitness function to select the candidate points that minimize the error of the calculation result as the initial value of the Gauss-Newton iteration method. The point with the highest fitness is selected as the initial value. The probability of selection is

$$P = \frac{fit(i)}{\sum_{j=1}^n fit(j)}. \quad (7)$$

The two-by-two connection of three nodes in this fractal design system can divide the points in the entire positioning area into seven ranges. Therefore, when selecting the initial value, one point in each of the seven regions can be randomly selected as a candidate point for the initial value. We use these seven points as the initial values in the Gauss-Newton iteration method to perform iterations and limit the number of iterations to 10. Within the specified number of times, the point where the fitness function takes the smallest value is taken as the final true initial value. This initial value is calculated iteratively, and a better initial value is selected for the iterative process. Each iteration corrects the calculated solution to obtain the optimal solution.

The final result obtained will be returned as the calculation result of target positioning. This ensures that the optimal result can be obtained conveniently in the locatable area. Figure 3 illustrates the calculation steps in detail.

## 4. Experiment and Discussion

**4.1. Optimization Results of Newton Iterative Algorithm.** The value of the Jacobian matrix obtained in the process of calculating the representative points of different regions is different, so the correction amount that determines the iteration is different. As a result, convergence and non-convergence occurred in the iterative calculation. The final result of iterative calculation of the initial values of different artistic graphic candidates is shown in Figure 4.

It can be seen from Figure 4 that the error of the calculation results of the initial values of different art graphics is less than 7%. This verifies that selecting the initial iterative value by region is effective in improving the traditional Gauss-Newton iteration method. Compared with the method of randomly selecting the initial value, selecting the initial value by region can improve the calculation accuracy of the Gauss-Newton iteration method. Most of the results obtained by the iterative process of randomly selecting initial values have large errors. The positioning result is largely limited by the distance between the randomly generated initial value position and the actual sound source. The smaller the distance between the randomly generated initial value of the iteration and the actual far sound point, the smaller the error of the calculation result and the more ideal calculation result. When the randomly selected initial value is far away from the actual sound source point, the calculation result has a large error. The positioning method of randomly generating initial values is limited to the range of generating random numbers, and, at the same time, due to its randomness, it is not suitable for practical systems.

**4.2. The Influence of Iterative Function and Parameter Changes on the Generation of Newton Iterative Graph.** The factors affecting the generation of Newton iterative graphics mainly include the type of the iterative function and the value of the parameters  $p$  and  $q$  and the selected Newton iterative graphics part in the design of fractal art graphics.

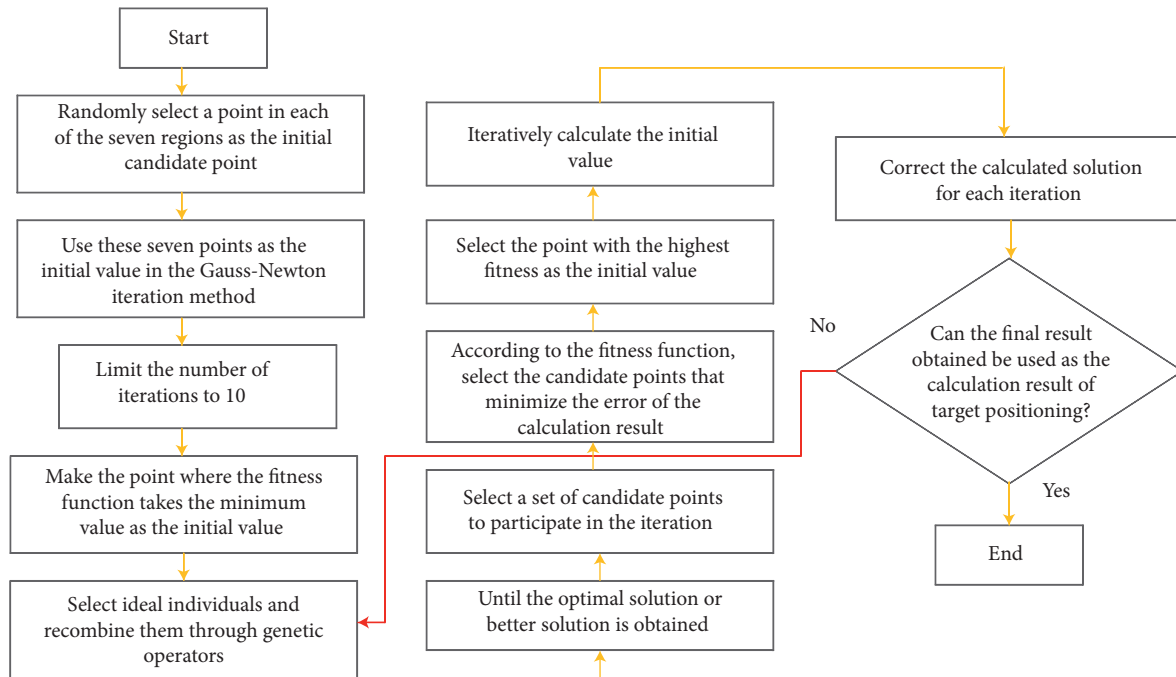


FIGURE 3: Flowchart of initial value selection.

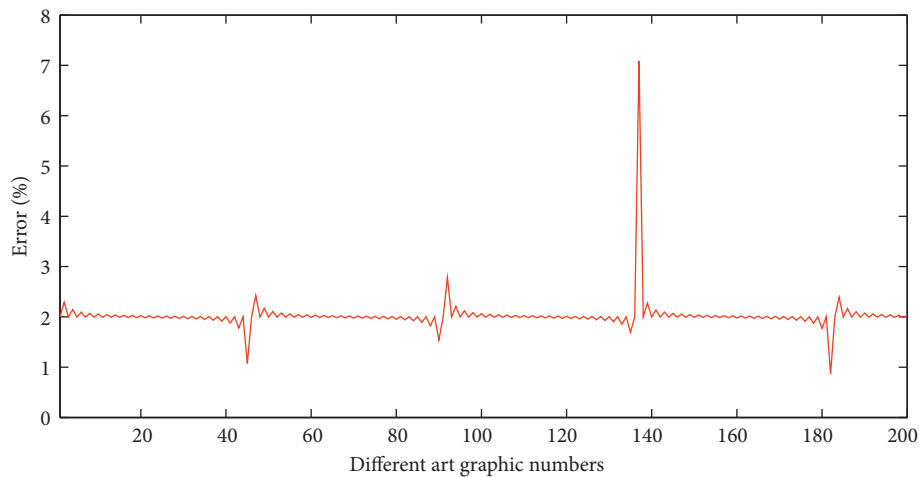


FIGURE 4: Calculation results of the initial values of different artistic graphics.

Through experiments and analysis and summary, some regular trends of Newton iterative graph changes can be grasped, and Newton iterative graphs with special texture effects can be generated.

According to the definition of Newton's iteration and its computer visualization principle, it can be known that the type of iteration function has a decisive effect on the formation of Newton's iterative graph fractal art. The iterative function can choose trigonometric function, power function, exponential function, hyperbolic function, absolute value function, and so forth. Selecting different iterative functions can get  $N$  set graphics with different shapes. Among them, trigonometric functions are divided into sine, cosine, tangent, cotangent, and other trigonometric

functions and can also include the power exponent change of the trigonometric function. This article mainly chooses trigonometric functions, power functions, exponential functions, and hyperbolic functions as the iterative functions of Newtonian iterative graphs. The Newton iteration graphs generated by different iteration functions are shown in Figure 5.

In the same type of iterative function, changing the iterative mapping function means changing the numerator and denominator of the real and imaginary parts of the iterative formula, and the reconstructed Newton iterative graph will also have a kaleidoscopic structure. This article has proposed a variety of methods to reflect the texture of Newton's iterative graphs in the design. Among them, the

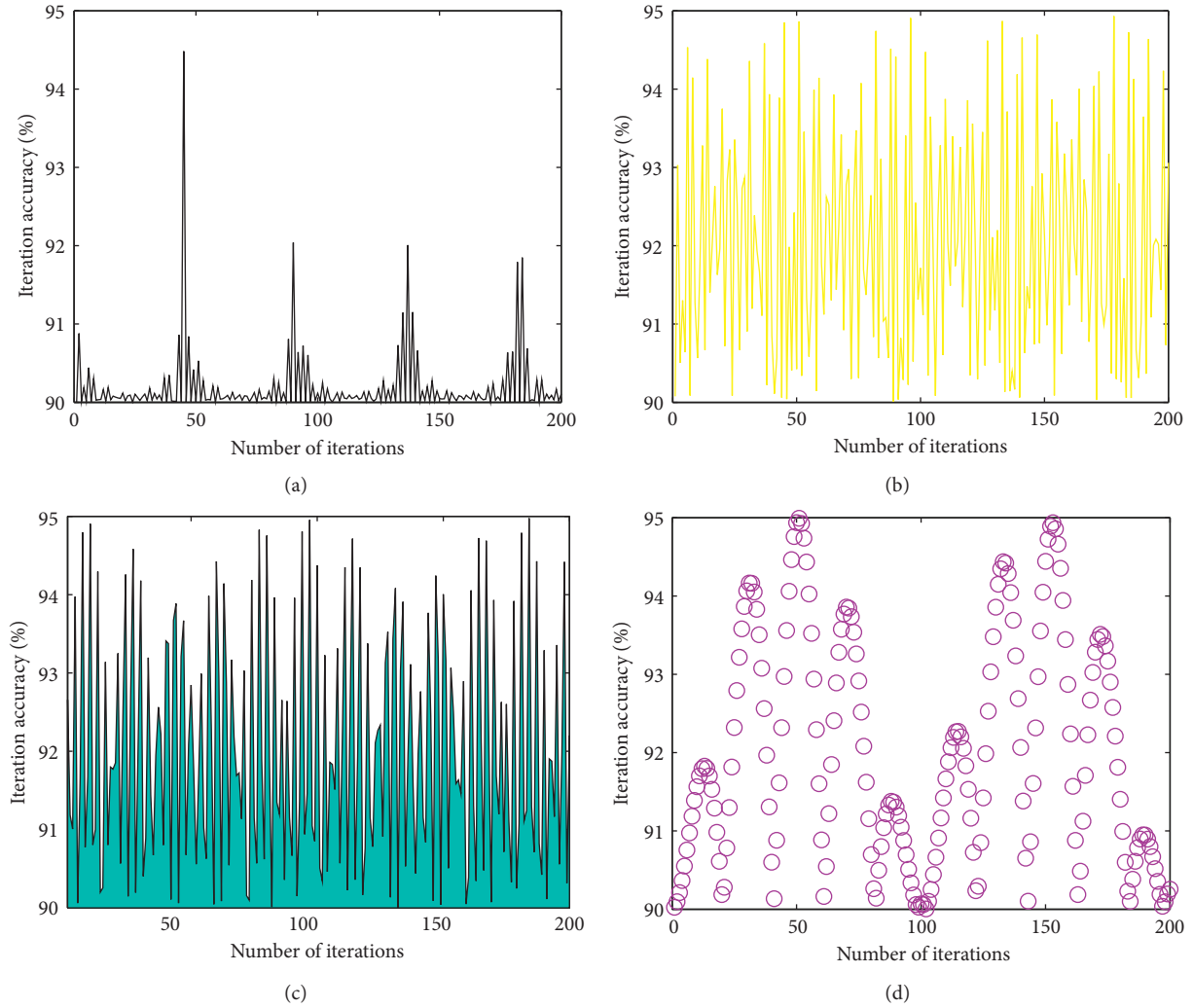


FIGURE 5: The accuracy of Newton iteration generated by different iteration functions. (a) Power function Newton iteration accuracy rate. (b) Trigonometric function Newton iteration accuracy rate. (c) Exponential function Newton iteration accuracy rate. (d) Hyperbolic function Newton iteration accuracy rate.

power function Newton's iterative graph and the trigonometric function Newton's iterative graph change greatly because of the change of the iterative function.

It is found through experiments that the power exponential Newton iteration graph and the trigonometric function Newton iteration graph are more sensitive to the changes of parameters, while the exponential function Newton iteration graph and the hyperbolic function Newton iteration graph have little adjustment to the graph structure caused by the change of the parameters.

*4.3. The Influence of Mathematical Morphology Image Processing Methods on Newton Iterative Graphs.* Through Matlab software, we have performed a variety of mathematical morphological processing on Newton's iterative graphs, including two operations, expansion and erosion. Figure 6 in this article is a sample of Newton's iterative graphics processed by mathematical morphology as fractal art graphics.

By comparison, it can be found that when the two structural elements of diamond and square are selected, the image effect after the corrosion operation is clearer, and the style of Newton's iterative graphics is more obvious. However, there are many block structures after image processing of structural elements, which will cause uneven tension in the warp and weft directions during design, which is not conducive to design.

*4.3.1. Selection of Structural Elements.* Choosing appropriate structural elements plays an important role in better expressing the special texture effects of Newton iterative graphics. Structural elements include spherical, linear, diamond, square, and disc. Through multiple experiments and comparisons, this article mainly uses two structural elements, diamond-shaped and square-shaped, in the experiment.

As shown in Figure 7, it is found through experiments that, with the increase of structural elements, Newton's

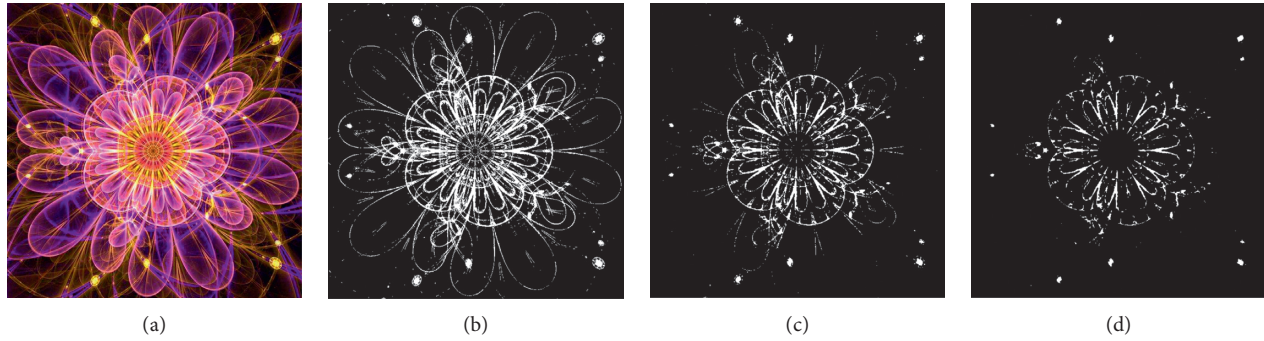


FIGURE 6: Comparison of Newton iteration graphs after corrosion calculation of different structural elements. (a) Original image. (b) Binary image after dither processing. (c) Corrosion of diamond structural element. (d) Corrosion of square structural element.

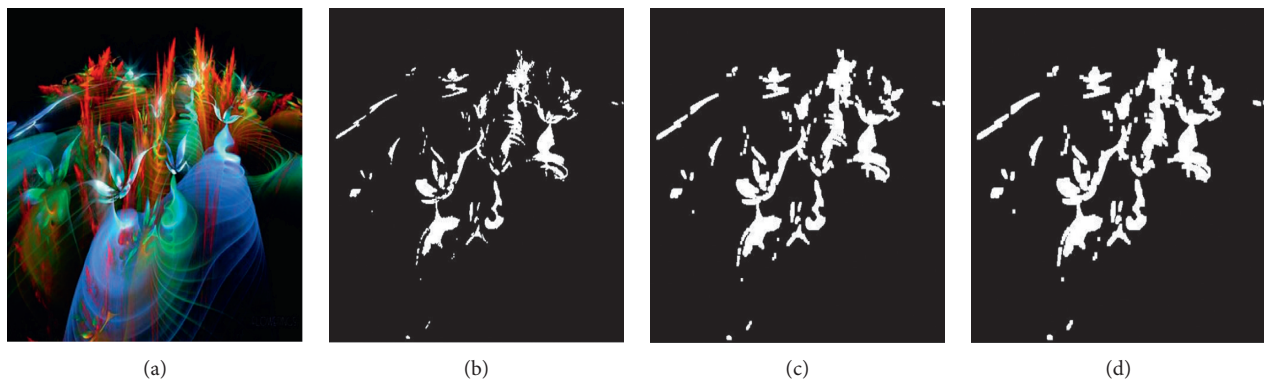


FIGURE 7: The Newton iteration graph after the expansion operation of the diamond structure element in different units. (a) Original image. (b) 1-unit diamond structural element expansion. (c) 2-unit diamond structural element expansion. (d) 3-unit diamond structural element expansion.

iterative graphics gradually turn from delicate to rough style, and their inherent texture effect gradually disappears.

**4.3.2. Application of Expansion Calculation.** The expansion operation is more suitable for nonscattered Newton iterative graphs. For scatter Newton iterative graphs, after the expansion operation processing, the scattered points will be combined with the scattered points of the attachment to become a block, and as the structural elements increase, the scattered points disappear quickly, and the unique texture effect also disappears.

**4.3.3. Application of Corrosion Calculation.** By comparison, it can be found that the erosion operation is more suitable for Newton iterative graphs of scatter points. When selecting the two structural elements of diamond and square, the image effect after the corrosion operation will be clearer than before, and the style of Newton's iterative graphics will be more obvious. With the increase of structural elements, the more blocky structures appear after image processing.

**4.4. The Influence of Fractal Art Graphic Organization and Fractal Art Graphic Density on Experimental Results.**

Traditional fractal art graphics are divided into flower parts and ground parts. Different colors are used to express in the weaving design, and the organization is corresponding; that is, one color corresponds to one organization. Newton iteration graphics are all composed of scattered points and thin lines, which determine that the flower parts are composed of scattered points and thin lines, and the rest are all ground parts, which need to be properly organized. In the experimental design of this paper, the same fractal art graphics are matched with 2–4 organizations, so that the warp and weft yarns of the fractal art graphics are more complex, the surface of the fractal art graphics reflects more details, and the texture is more complex and delicate.

The fractal art graphics used in this topic are inherently complex and delicate, which require delicate materials, better gloss, and the highest possible density of yarns. The density of cotton fractal art patterns can reach up to 70–85 pieces/cm, and the density of silk fractal art patterns can reach up to 190 pieces/cm. Therefore, it can be seen that real silk has a great advantage in reflecting the delicate structure of fractal art graphics. For this reason, we choose real silk as raw material, and we can also try to use cotton yarn, chenille, polyester, and other raw materials to obtain different style effects. The fractal accuracy rate of art graphics of the improved Newton iterative algorithm is shown in Figure 8.

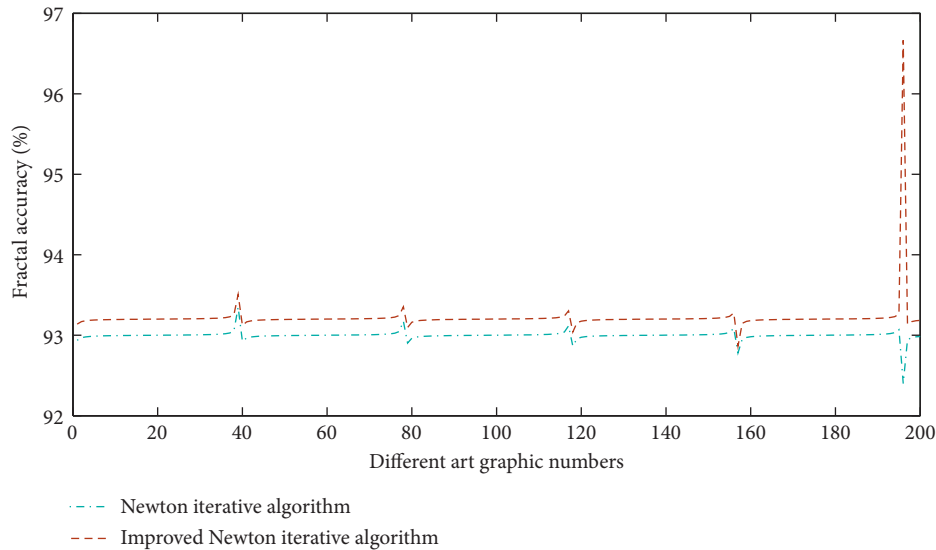


FIGURE 8: The fractal accuracy of artistic graphics of the Newton iterative algorithm before and after the improvement.

## 5. Conclusion

This paper analyzes the basic principles of the Gauss-Newton iteration method and we found that the Gauss-Newton iteration method is greatly affected by the initial value of the iteration. If the initial value is not properly selected, the iteration may not converge to the wrong result. Conversely, selecting appropriate initial values can efficiently calculate accurate results and reduce positioning errors. A representative point is selected from the seven regions as the candidate initial value and substituted into the equation for iterative calculation. We use the value with the smallest error as the final calculation result. The method of selecting the initial value in the Gauss-Newton iteration method is optimized. We construct a new fractal art graphic design and form model based on Newton's iterative theory. Specifically, it includes changing various factors that affect the generation of Newton iterative graphs, continuously transforming the factors that affect the generation of Newton iterative graphs and summarizing the regularity of their changes. In order to find a special type of Newton iterative graphics, we design fractal art graphics and then select the appropriate organization to reflect the unique mechanism of Newton's iterative graphics. We use MATLAB to perform morphological transformation on the designed fractal art graphics to obtain the transformed fractal art graphics. We use this kind of fractal art graphics to design fractal art graphics with special texture effects with the help of weaving CAD and Jacquard design technology. Through factors such as the selection of fractal art graphic organization and the change of the size of the fractal art graphic cycle, we design fractal art graphic textures with different effects.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no known conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

## References

- [1] K. Gdawiec, A. A. Shahid, and W. Nazeer, "Higher order methods of the basic family of iterations via S-iteration scheme with s-convexity," *Mediterranean Journal of Mathematics*, vol. 17, no. 2, pp. 1–19, 2020.
- [2] S. Kumari, M. Kumari, and R. Chugh, "Dynamics of superior fractals via Jungck SP orbit with s-convexity," *Annals of the University of Craiova-Mathematics and Computer Science Series*, vol. 46, no. 2, pp. 344–365, 2019.
- [3] I. Gościński and K. Gdawiec, "Control of dynamics of the modified Newton-Raphson algorithm," *Communications in Nonlinear Science and Numerical Simulation*, vol. 67, pp. 76–99, 2019.
- [4] H. Ahmad, A. R. Seadawy, T. A. Khan, and P. Thounthong, "Analytic approximate solutions for some nonlinear parabolic dynamical wave equations," *Journal of Taibah University for Science*, vol. 14, no. 1, pp. 346–358, 2020.
- [5] M. Bisheh-Niasar and K. Gdawiec, "Bisheh-Niasar-Saadatmandi root finding method via the S-iteration with periodic parameters and its polynomiography," *Mathematics and Computers in Simulation*, vol. 160, pp. 1–12, 2019.
- [6] N. H. Khan, Y. Wang, D. Tian, M. A. Z. Raja, R. Jamal, and Y. Muhammad, "Design of fractional particle swarm optimization gravitational search algorithm for optimal reactive power dispatch problems," *IEEE Access*, vol. 8, pp. 146785–146806, 2020.
- [7] S. Hasan, A. Al-Zoubi, A. Freihet, M. Al-Smad, and S. Momani, "Solution of fractional SIR epidemic model using residual power series method," *Applied Mathematics & Information Sciences*, vol. 13, no. 2, pp. 153–161, 2019.
- [8] S. Ahmed, K. K. Ghosh, P. K. Singh, Z. W. Geem, and R. Sarkar, "Hybrid of harmony search algorithm and ring theory-based evolutionary algorithm for feature selection," *IEEE Access*, vol. 8, pp. 102629–102645, 2020.



- [9] J. Colangelo, "The mote in god's i: fourier, fractals, and imaginary numbers in tom stoppard's arcadia," *Modern Drama*, vol. 62, no. 2, pp. 191–207, 2019.
- [10] O. Acan, O. Firat, and Y. Keskin, "Conformable variational iteration method, conformable fractional reduced differential transform method and conformable homotopy analysis method for non-linear fractional partial differential equations," *Waves in Random and Complex Media*, vol. 30, no. 2, pp. 250–268, 2020.
- [11] R. Fischer, P. Dittmann, R. Weller, and G. Zachmann, "AutoBiomes: procedural generation of multi-biome landscapes," *The Visual Computer*, vol. 36, no. 10–12, pp. 2263–2272, 2020.
- [12] H. Mittal and M. Saraswat, "An automatic nuclei segmentation method using intelligent gravitational search algorithm based superpixel clustering," *Swarm and Evolutionary Computation*, vol. 45, pp. 15–32, 2019.
- [13] S. K. Katiyar, A. K. B. Chand, and G. Saravana Kumar, "A new class of rational cubic spline fractal interpolation function and its constrained aspects," *Applied Mathematics and Computation*, vol. 346, pp. 319–335, 2019.
- [14] G. Singh, B. Singh, and M. Kaur, "Grasshopper optimization algorithm-based approach for the optimization of ensemble classifier and feature selection to classify epileptic EEG signals," *Medical & Biological Engineering & Computing*, vol. 57, no. 6, pp. 1323–1339, 2019.
- [15] A. A. Ansari, R. Kellil, Z. A. Al-Hussain, and W. Ul-Haq, "Effect of variation of charge in the circular restricted three-body problem with variable masses," *Journal of Taibah University for Science*, vol. 13, no. 1, pp. 670–677, 2019.
- [16] L. D. Petković, M. S. Petković, and B. Neta, "On optimal parameter of Laguerre's family of zero-finding methods," *International Journal of Computer Mathematics*, vol. 96, no. 4, pp. 692–707, 2019.
- [17] M. Umar, Z. Sabir, and M. A. Z. Raja, "Intelligent computing for numerical treatment of nonlinear prey-predator models," *Applied Soft Computing*, vol. 80, pp. 506–524, 2019.
- [18] O. P. Verma, G. Manik, and S. K. Sethi, "A comprehensive review of renewable energy source on energy optimization of black liquor in MSE using steady and dynamic state modeling, simulation and control," *Renewable and Sustainable Energy Reviews*, vol. 100, pp. 90–109, 2019.
- [19] T. Murugesan, S. Varshini, and M. D. Kumar, "Homotopy perturbation method of hydromagnetic flow and heat transfer of a casson fluid over an exponentially shrinking sheet," *World Scientific News*, vol. 140, pp. 59–78, 2020.
- [20] M. A. Tawhid and A. M. Ibrahim, "A hybridization of grey wolf optimizer and differential evolution for solving nonlinear systems," *Evolving Systems*, vol. 11, no. 1, pp. 65–87, 2020.

## Research Article

# Resource Scheduling and Strategic Management of Smart Cities under the Background of Digital Economy

Qing Yin <sup>1,2</sup> and Gang Liu<sup>1</sup>

<sup>1</sup>School of Business, East China University of Science and Technology, Xuhui, Shanghai 200237, China

<sup>2</sup>Henan University of Animal Husbandry and Economy, Zhengzhou, Henan 450000, China

Correspondence should be addressed to Qing Yin; [y10180446@mail.ecust.edu.cn](mailto:y10180446@mail.ecust.edu.cn)

Received 3 October 2020; Revised 13 November 2020; Accepted 17 November 2020; Published 28 November 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Qing Yin and Gang Liu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Smart city is a brand-new city form, in which information and communication technologies are utilized to sense, analyze, and integrate the key information of city operation core system, so that intelligent responses can be immediately and effectively taken to various demands including people's livelihood, environmental protection, public safety, city services, and industrial and commercial activities. Digital economy is a mixed economy with the coexistence of multiple business models and diversified value creation models based on the information and communication technologies and in the digital economy, many things are undergoing huge changes, and their corresponding economic rules also need to be adjusted. On the basis of analyzing previous research works, this paper expounded the research status and significance of smart city's resource scheduling and strategic management, elaborated the development background, current status, and future challenges of digital economy, introduced the methods and principles of city-level spatiotemporal data model and spatial full factor coding, formulated resource scheduling strategies for smart city based on digital economy, explored the dynamic fusion, storage, and update of smart city's multisource heterogeneous data, conducted the information display and analysis of multilevel smart city, proposed strategic management approaches for smart city based on digital economy, analyzed the integrated implementation model of shared resource scheduling and people-oriented social management, and discussed the economic growth factors and standardization mechanism of smart city under the background of digital economy. The results of this study provide a reference for further research studies on the resource scheduling and strategic management of smart city under the background of digital economy.

## 1. Introduction

Smart city is a brand new city form which builds a smart and intelligent environment to support the city development; it utilizes cutting-edge information technology methods, such as the Internet of Things, cloud computing, optical networks, and mobile Internet, to integrate the scattered and fragmented information systems in a city into an organic whole with better synergy and regulation capabilities, so as to intelligently respond to various needs of public services, social management, industrial operations, and other activities [1]. The essence of a smart city is to make use of advanced information technology to realize smart management and operation of the city, thereby creating a better life for people in the city and promoting the

harmonious and sustainable growth of the city [2]. The security of a smart city is a key indicator of its information-awareness layer. The analysis of city security characteristics is the basis for the construction of a smart city model, but this analysis model does not take into account factors such as the city environment, government services, and infrastructure. The construction of a new type of smart city includes four key elements of the open architecture of the Internet of Things, the city's open information platform, the city operation command center, and cyberspace security. The development of the concept of smart city and related applications can promote the realization of a consumption-driven economic development mode from the three aspects of increasing household consumption, corporate consumption, and government consumption [3]. In addition,

the information, the Internet of Things, and related industry development brought by the development of smart city can also promote industrial structure adjustment from the three aspects of coordinating the proportion of the three industries, improving industrial quality, and developing emerging industries and these three aspects reflect to varying degrees rationalization and advanced industrial structure [4].

Digital economy is a mixed economic system, in which information technology and computer technology are widely used. The use of information technology and computer technology has fundamentally changed the mode and situation of economic environment and economic activities [5]. The management system and multiple departments are indeed very important for the overall deployment and construction planning tasks of a smart city. In the digital economy, the transactions of management system usually have the characteristics of digitization, virtualization, concealment, and electronic payment in its multiple departments, which make it difficult to judge the consumption of trading venues and provide goods and services. In addition, the digital economy can provide better support for production, distribution, sales, and other goods and services, thereby creating a win-win environment for businesses and consumers [6]. The essence of the digital economy lies in informatization, which is a completely new social and political and economic system that fully digitizes information and business activities, including the development of information technology such as information technology, traditional industries, infrastructure, and social life [7]. The era of the digital economy is an era of globalized market competition, knowledge, and convergence and knowledge is spreading, accumulating, and updating more quickly due to the development of information technology and communication networks. The smart city is the main carrier of the development of the digital economy; digital economy is also the main feature of the development of the industrial economy of the smart city and the two are mutually reinforcing and mutually dependent relationships [8].

On the basis of analyzing previous research works, this paper expounded the research status and significance of smart city's resource scheduling and strategic management, elaborated the development background, current status, and future challenges of digital economy, introduced the methods and principles of city-level spatiotemporal data model and spatial full factor coding, formulated resource scheduling strategies for smart city based on digital economy, explored the dynamic fusion, storage, and update of smart city's multisource heterogeneous data, conducted the information display and analysis of multilevel smart city, proposed strategic management approaches for smart city based on digital economy, analyzed the integrated implementation model of shared resource scheduling and people-oriented social management, and discussed the economic growth factors and standardization mechanism of smart city under the background of digital economy. The detailed section arrangement is as follows: Section 2 introduces the methods and principles of city-level spatiotemporal data model and spatial full factor coding; Section 3 formulates resource scheduling strategies for smart city based on digital

economy; Section 4 proposes the strategic management approaches for smart city based on digital economy; Section 5 discusses the economic growth factors and standardization mechanism of smart city; Section 6 is conclusion.

## 2. Methods and Principles

*2.1. City-Level Comprehensive Spatiotemporal Data Model.* Due to the different dimensions of the indicators in the evaluation index system for digital economy development, the orders of magnitude differ greatly and they tend to increase [9]. Therefore, the initial data of the basic data is transformed into a comparable sequence and the calculation method is as follows:

$$P_{ij} = \frac{X_{ij}}{Y_{ij}}, \quad (1)$$

where  $P_{ij}$  is the data of the  $i$ -th index at the beginning of the  $j$ -th year;  $X_{ij}$  is the original data; and  $Y_{ij}$  is the value of the  $i$ -th index at the  $j$ -th year and the weight of each index is calculated as follows:

$$S_{ij} = -k \cdot \sum_{i,j=1}^n p_{ij} \ln p_{ij}, \quad i = 1, 2, \dots, n, \quad (2)$$

where  $k$  is the fractal coefficient of the digital economy;  $p_{ij}$  is the proportion of the value of the  $i$ -th index in the  $j$ -year value of the indicator; and  $S_{ij}$  is the information entropy of the  $i$ -th index in the  $j$ -th year.

According to the characteristics and basic structure of smart city, the intelligent system of city public facilities based on digital economy is designed and the model of building smart buildings can be expressed as

$$S_{ab} = \frac{\alpha(a_1 - b_2) + \beta(a_1 + b_2)}{S_a^\beta \cdot S_b^\alpha} + X_{ab} \cdot Y_{ab}, \quad (3)$$

where  $a_1$  and  $b_2$  are the factors of the digital economy system;  $\alpha$  is the feature space factor of the smart city;  $\beta$  is the dimension of the divergence matrix;  $S_a^\beta$  and  $S_b^\alpha$  are the inter-class divergence matrix and intra-class divergence matrix, respectively; and  $X_{ab}$  and  $Y_{ab}$  are the city dimension spatial data sample set and feature attribute points, respectively.

Constraint equations that use the digital economy to optimize city-level spatiotemporal data are

$$W(S_{ij}) = \begin{cases} w_1, & 0 < S_{ij} \leq h_1, \\ w_2, & d_1 < S_{ij} \leq h_2, \\ \vdots & \vdots \\ w_k, & h_{k-1} < S_{ij} \leq 1, \end{cases} \quad (4)$$

where  $W(S_{ij})$  is the constraint condition that the average dimension of the parameters of smart optimization system needs to meet;  $w_k$  is the interference caused by the digital economy background to city construction problems; and  $h_k$  is the global information feature vector of the smart city. So the city-level spatiotemporal comprehensive data model is calculated as

$$Q_x = P_{ij} \cdot W(S_{ij}) + \xi \cdot \frac{f(X_{ij}, Y_{ij}) \cdot S_{ab}}{\gamma \cdot (x_i + y_j)}, \quad (5)$$

where  $Q_x$  is the equipment mark of the digital economy;  $\xi$  is the correlation factor;  $f(x_{ij}, y_{ij})$  represents the regression basis function;  $x_i$  is the principle of minimizing the intelligent use of digital resources;  $y_j$  is the city optimization factor; and  $\gamma$  is the minimum value of the constraint.

**2.2. City-Level Spatial Full Factor Coding.** The digital economy's impact on the current tax system is comprehensive, involving multiple aspects of the impact on the elements of the tax system itself, the division of tax jurisdiction, and the impact on the current tax collection and management system [10]. When the digital economy shows extremely complex shapes, it cannot be expressed by simple mathematical expressions, but only approximate mathematical functions can be used to approximate the actual situation. According to its coding requirements for city spatial elements, its fitting function  $P(x_i, y_i, z_i)$  can be expressed as

$$P(x_i, y_i, z_i) = \sum_{i=1}^n S_i \left( \frac{x_i^2}{Q_x^2} \cdot \ln \frac{y_i^2}{Q_x^2} + \frac{z_i^2}{Q_x^2} \right). \quad (6)$$

In the formula,  $Q_x^2$  is the influence radius of discrete points;  $Q_x^2 = (x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2$ ;  $x_i$ ,  $y_i$ , and  $z_i$  ( $i = 1, 2, \dots, n$ ) are the sampling data of discrete points on the digital system; and  $S_i$  is the undetermined coefficient, given by the information of the known points.

The city-level spatial all-elements group is updated synchronously; that is, after  $m$  elements complete their respective traversal searches, the digital information is updated and adjusted uniformly. If  $k_{ij}(n)$  is assumed as the information factor after the  $n$ th cycle, the following formula can be obtained:

$$k_{ij}(n) = (1 - \lambda)k_{ij}(n-1) + \sum_{i,j=1}^m \Delta k_{ij}^n, \quad (7)$$

where  $\lambda$  is the volatility coefficient of the total space factor,  $(1-\lambda)$  is the digital information, and it is also the pheromone residual factor, and  $\Delta k_{ij}^n$  is the digital increment of the  $k$ -th space full factor on the path in this cycle.

In order to calculate the coding time required for each spatial element of a smart city under different digital economic conditions, the execution matrix is defined as

$$H(i, j) = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1m} \\ h_{21} & h_{22} & \cdots & h_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ h_{n1} & h_{n1} & \cdots & h_{nm} \end{bmatrix}, \quad (8)$$

where  $k_{ij}$  represents the encoding time required to represent the digital economic conditions  $h_i$  to process the full space elements  $h_j$ . If  $H(i, j)$  is assumed as the gradient value of the previous feature point, and  $J(i, j)$  is the gradient value of the

neighboring features centered on the current feature point, the gradient discrimination factor  $R(i, j)$  can be defined as

$$R(i, j) = \max\{|J(i, j)| - |H(i, j) \cdot J(i, j)|\}. \quad (9)$$

It can be known from the above formula that the gradient discrimination factor  $R(i, j)$  is the minimum value of the difference between the gradient modulus value of the current element and the  $n$  neighboring point modulus values in the digital economic model; then, the adaptive fractional order expression is defined as

$$A(i, j) = \begin{cases} \frac{J(i, j) - H(i, j)}{\max\{J(i, j)\} - \nu \cdot R(i, j)}, & R(i, j) \geq K, \\ 0, & R(i, j) < K. \end{cases} \quad (10)$$

In the formula,  $\nu$  is the adjustment coefficient of spatial full-element coding;  $\max\{G(i, j)\}$  is the maximum gradient of  $n$  feature points; and  $K$  is the threshold value selected according to different smart city models. When the required gradient value of the processed space and the gradient value of the surrounding elements change greatly, that is, when the gradient discrimination factor is greater than a given threshold value, it is considered that the encoding of the spatial full factor is successful.

### 3. Resource Scheduling of Smart City Based on Digital Economy

**3.1. Dynamic Fusion, Storage, and Update of Multisource Heterogeneous Data.** The spatial and temporal characteristics are the obvious characteristics of each element in the city. The information analysis of the smart city should straighten out the relationship between different elements, find similarities among them, and centrally process the information of all parties. The city-level spatiotemporal comprehensive data model can not only complete the establishment of a city-level spatiotemporal comprehensive data conceptual model from the perspectives of multiple dimensions, multiple temporalities, multiple dimensions, or even the four dimensions of multiple topics. The intangibility and mobility of the digital economy make the connection factors that determine tax jurisdiction disappear or become obscured, which brings challenges to the traditional standards for determining tax jurisdictions of residential places and source jurisdictions. In the application of comprehensive information, comprehensive analysis of multisource heterogeneous business, and integration sorting, processing and clarification of data are analyzed to determine the dwelling form of information collection and construction of judgment paths to provide multitopic information service. The framework of resource scheduling of smart city under the background of digital economy is shown in Figure 1. The information system design indeed needs to be continuously adjusted and updated as the environment changes. The digital economy mainly presents the economic process of manufacturing and supplying electronic products and services for commerce or trading activities caused by commerce. In this process, production and

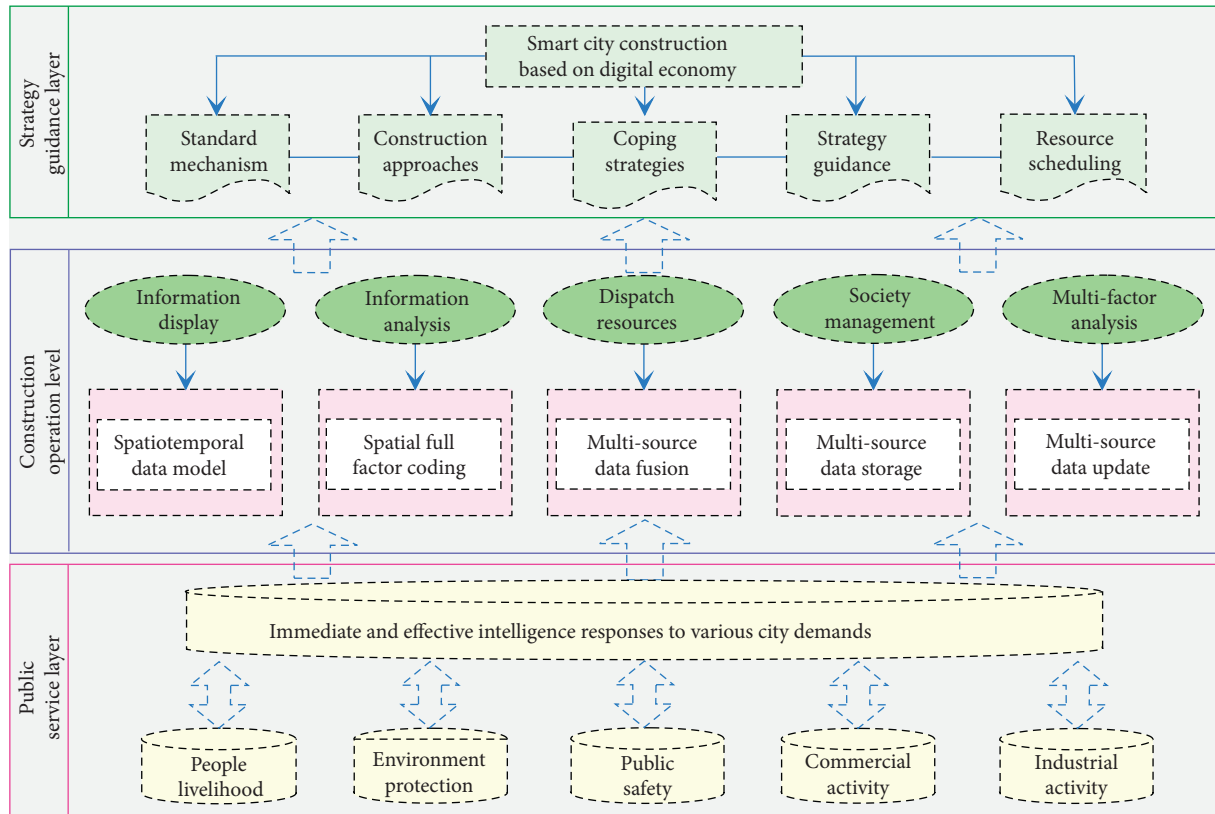


FIGURE 1: Framework of resource scheduling of smart city under the background of digital economy.

management are electronic as the environment changes, in which various suppliers and consumers and related activities are connected through the information analysis technology.

A smart city is of comprehensive structural design, technology integration, and three-dimensional services. Technical cooperation in the areas of city governance, city planning, industrial informatization, and architecture is required to achieve a high degree of integration, resource optimization, and highly interconnected technology and functions. At the same time, it is also a strategy for advancing smart city, which is different from the core of the previous informatization and digital cities. It also emphasizes the organic connectivity and functional realization of city households between different industries, internal industries, and other factors. Therefore, the coordination of the smart city industry has become a key path to accelerate the landing of the smart city industry. Most of the key projects in the areas of smart life, smart city management, and smart government affairs are invested and led by the government. The starting point is mainly from serving city management and administrative approval and there is a certain misalignment between project design, business services, and citizen demand. From supply to demand, the model is usually not timely, accurate, and comprehensive to meet the needs of citizens, and the government has not established an information-based active service mechanism for the needs of social entities. Smart city construction funds are mainly used in transportation, security, and other fields that have an important impact on city construction and management,

not at the community level, which is closely related to people's livelihood; social capital investment has not formed an effective business model [11].

In order to complete the overall deployment and construction planning tasks of a smart city, the management system must be first improved and multiple departments should be involved in the implementation of the planning and design process. Therefore, a dedicated smart city department must be established to form the smart city's affairs and materials. The management mechanism of talents and people is followed by effective planning and top-level design of a smart city to effectively adapt to modern development, and then the construction of a city performance evaluation system, to effectively ensure the healthy development of the city [12]. In order to enhance the driving force of smart city planning and design, it must actively implement innovation, rely on the latest science and technology to achieve effective docking with smart city, and ensure the source of power for smart city planning and design. The technological innovation environment must be first optimized and the public integrated process of the service platform then should be improved to guide the whole society to actively innovate and start businesses. The smart home function can be added, and city residents can centrally control the homes at home and they can also understand the price and quality of different home products in a unified interface. At the same time, they can set up corresponding merchant connection services to ensure that merchants can integrate the latest home products and push residents to build a contact platform for merchants and residents.

*3.2. Multilevel City Comprehensive Information Display and Analysis.* An information system cannot be successfully designed at one time, which is necessary to improve the design incompletely during operation and continuously adjust and update as the environment changes and the self-improving ability is the core of the system's vitality. The long-acting system is not a static system, but an adaptive system with a self-regulating improvement mechanism, and the project management mechanism is the key point. Choosing an appropriate management model is a guarantee for the long-term benefits of the project and projects that do not fully consider the management model are not feasible. The selectable management mode of the city's intelligent system which includes government, cause and non-profit organization and enterprise. The different management modes and mechanisms have their own advantages and should be selected according to the characteristics of the service business [6]. The key to the sustainability of an information sharing system is to establish a quality assurance mechanism for data resource providers. Innovative ideas need to be reflected through project optimization, and planning attention should be focused on policy design and project selection. Table 1 shows the rating system of smart city resource scheduling based on digital economy. The smart city is not expected to bring shortcuts to information construction and information is only a tool for city development and the benefits of the tools need to be matched with the environment. The transformation and reconstruction of the environment must not be abrupt, and the smart city cannot bring sudden changes; mentality does not always want to leapfrog development and there are many benefits achieved by reducing the scale of the project.

The core of a smart city is a smarter way to change the way governments, communities, or companies and people interact with each other through a new generation of information technology to improve the clarity, efficiency, flexibility, and responsiveness of interactions with basic characteristics of perception, interconnection, and intelligence. From a technical perspective, the basic structure of a smart city mainly includes the Internet of Things, data center, city management, and operation levels, which correspond to the smart city intelligent sensing network, information resource center, operation management application layer, and public service center. Figure 2 shows the relationship between encoding relative time and index system rating in the resource scheduling and strategic management of smart city. The core technologies of smart city include intelligent identification, mobile computing, cloud computing, and information fusion and the public information services usually refer to the process of implementing open and developing services for public information. It includes two levels of open services and development services, including information on public affairs management activities, as well as information related to public interests, public policies, and arrangements and implementation of public management systems. Under the strategic guidance of the top-level design, starting from the core function of public information service push, the innovation of the value supply of the platform and the business model is

achieved, driving the innovation of smart city form through the integration of multistakeholders.

The construction of a smart city is a complex and systematic project and building specific smart projects to meet the objective needs of the city is the basic principle for completing the construction. The existing city characteristic analysis models analyze different cities by using population, economic scale, traffic conditions, industrial strength, service industry strength, and innovation capacity as analysis factors. Among them, the population factor is calculated using the number of city populations; the economic scale is calculated using the regional gross domestic product; the traffic conditions are calculated using the turnover of goods or the turnover of tourists; the authorization calculation of industrial strength is calculated using the added value of the industry or the added value of the secondary industry. The analysis of city characteristics is the basis for the construction of a smart city model, but this analysis model does not take into account factors such as the city environment, government services, infrastructure, etc. The resulting analysis results may not meet the needs of actual work and there are certain limitations in a city which can be regarded as an organization, and a smart city is the result of the development of city informatization to a certain stage. To this end, the theoretical model applicable to the development of organizational informatization can theoretically also be used to build a smart city development path and city informatization. By analyzing the stage of each city's informatization, combined with the city's development goals and needs, it can better build the development path of smart city [13].

## **4. Strategic Management of Smart City Based on Digital Economy**

*4.1. Integrated Management of Sharing Scheduling City Resources.* Market demand is the driving force behind the sustainable development of smart city. The focus of smart city construction is to give play to the role of the market in allocating resources, to create various personalized smart applications that meet market needs through price, competition, and other factors, to foster a good state of emerging markets, and to promote new growth in the smart city economy. Technological innovation and the development of strategic emerging industries are important foundations for the construction of smart city. First of all, the construction of public service platforms must be strengthened in technology research and development, application testing, evaluation, and inspection, and the technological innovation environment also needs to be improved (Figure 3). The second is to promote the transformation of the dynamic mechanism of city development, improve the city industrial structure, and promote the emergence of new industrial forms with smart technological innovation. Finally, the key to strengthening technological research and development is to cultivate professional talents and make them the intellectual resources for smart city development. The construction process needs to make a reasonable top-level design based on the nature, characteristics, and functions of the city and establish a long-

TABLE 1: Rating system of smart city resource scheduling based on digital economy.

| Rating index system          | Smart foundation operation | Smart management services | Smart economy humanity | Smart integrated support |
|------------------------------|----------------------------|---------------------------|------------------------|--------------------------|
| Network connection level     | 0.836                      | 0.623                     | 0.735                  | 0.838                    |
| Service interaction platform | 0.783                      | 0.416                     | 0.511                  | 0.769                    |
| Technology infrastructure    | 0.804                      | 0.327                     | 0.850                  | 0.642                    |
| City marking technology      | 0.735                      | 0.020                     | 0.255                  | 0.523                    |
| Operation management         | 0.843                      | 0.134                     | 0.146                  | 0.114                    |
| Personal consumption         | 0.632                      | 0.463                     | 0.347                  | 0.745                    |
| Society environment          | 0.957                      | 0.032                     | 0.966                  | 0.758                    |
| Culture environment          | 0.362                      | 0.217                     | 0.675                  | 0.483                    |
| Government policy            | 0.235                      | 0.948                     | 0.107                  | 0.956                    |
| Business application         | 0.145                      | 0.566                     | 0.524                  | 0.046                    |
| Business environment         | 0.521                      | 0.211                     | 0.122                  | 0.919                    |
| Analysis function            | 0.165                      | 0.925                     | 0.694                  | 0.544                    |
| Cloud management             | 0.238                      | 0.866                     | 0.437                  | 0.312                    |
| Monitoring warning           | 0.366                      | 0.618                     | 0.724                  | 0.073                    |
| Resource scheduling          | 0.122                      | 0.345                     | 0.159                  | 0.479                    |
| Standard system              | 0.434                      | 0.584                     | 0.335                  | 0.249                    |
| Information security         | 0.593                      | 0.483                     | 0.135                  | 0.185                    |
| Decision support             | 0.453                      | 0.293                     | 0.674                  | 0.835                    |
| Solution logic               | 0.065                      | 0.484                     | 0.484                  | 0.686                    |
| Strategy guidance            | 0.968                      | 0.204                     | 0.597                  | 0.548                    |

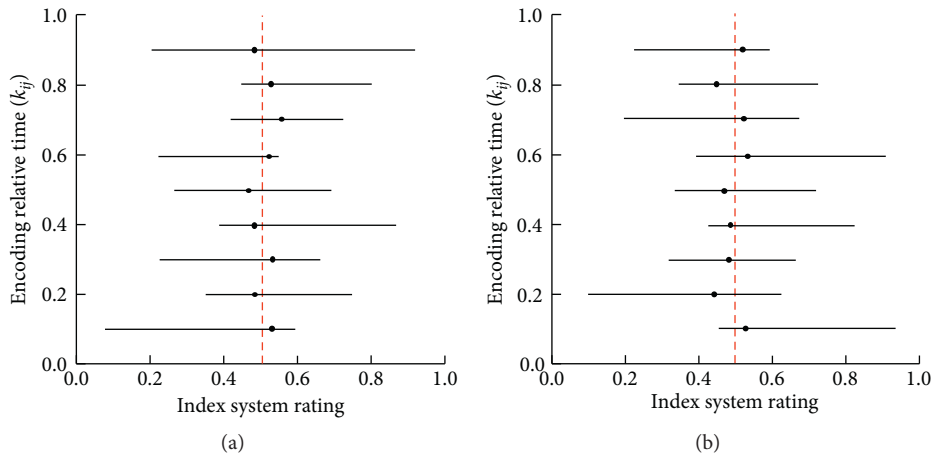


FIGURE 2: Relationship between encoding relative time and index system rating in the (a) resource scheduling and (b) strategic management of smart city based on digital economy.

term institutional guarantee. At present, the construction of smart city is mainly to give full play to advantages or solve shortcomings. Depending on the focus of city construction, the common goal is to identify obstacles to their own development; in some cities, the focus of construction is to give play to their advantages and promote economic development in a targeted manner.

A smart city is a complex large-scale system project, and its security problems will manifest in the system's information-awareness layer, access and transmission layer, application layer and terminal layer, intelligent processing, and collaborative platform layer. If sudden safety problems are encountered, the impact will involve the entire city, which may lead to factory shutdowns, store shutdowns,

electrical interruptions, traffic paralysis, water cuts, etc., which will have extremely serious consequences [14]. Therefore, they should be practically paid more attention to and dealt with and its heterogeneous interconnection, multihop, and distribution characteristics will make it more difficult to integrate and integrate its security system due to the use of a heterogeneous system platform. The different methods of equipment, storage, processing, and detection make information security, information transmission, and processing, which are also difficult to unify; the equipment is often in unattended, adaptive management, self-disconnection, and self-connecting states, which also increases the difficulty of designing and implementing security systems. Specifically, the integrity, authenticity, confidentiality,

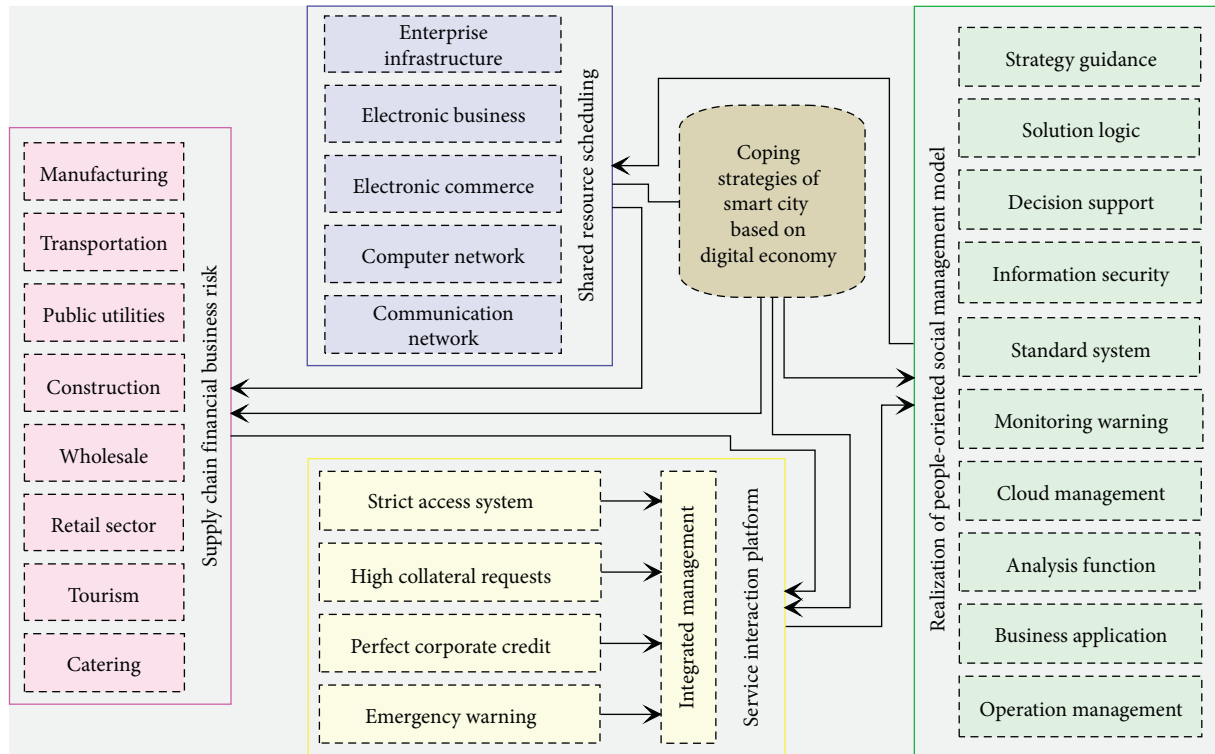


FIGURE 3: Integrated and people-oriented strategic management of smart city based on digital economy.

availability, fault tolerance, and self-organizing ability of massive data collection should be guaranteed for the ubiquitous network sensing layer with self-organizing characteristics. For this reason, special attention should be paid to solving physical-level attacks for obtaining stored passwords and data, and to handle encryption, decryption, identity authentication and signature, and key security issues.

In the future, the data of smart city will be aggregated and dispersed and convergence means that public information platforms gather, extract, and analyze various heterogeneous and heterogeneous data. The dispersion means that data from vertical industries or different government agencies and enterprises will be separated due to their natural nature which may be distributed across different private and public cloud data centers. According to the city's budget and planning, the solution can be based on three key platforms: infrastructure platform, centralized management, and flexible scheduling of network resources, allowing more embedded infrastructure. The intelligent processing platforms can consolidate and improve service delivery, intelligent innovation platform, and modular service channels, including user self-service, social networking sites, application incubation, and service support call centers by identifying service priorities, expanding [15]. The key business capabilities of smart city should move closer to the trends of cloud computing and mobile and social networks. As cloud computing deployments and the number of users increase, vendors will inevitably transition to service providers or choose the right partners. The social networks can not only capture and analyze user data in real time, but also

provide important feedback mechanisms that the public can use socially software or feedback to the government and manufacturers on services and suggestions at any time through social media, promote the improvement of service quality, and create more business opportunities.

#### 4.2. Realization of People-Oriented Social Management Model.

The smart city is supported by spatiotemporal information and relies on modern information technologies such as the Internet of Things and cloud computing to materialize human knowledge into various activities such as city planning, design, construction, management, operation, and development under informative conditions. The industry sector proportion of digital economy in smart city construction is shown in Table 2. A smart city should include real-time information sensing devices like human senses, two-way transmission of information and instructions like human nervous systems, ubiquitous network systems, and cloud computing service centers like human brains, and response and human behavior organs. A smart city cannot be separated from the support of the geospatial framework [16]. The geospatial framework is the responsibility of the geographic information department of mapping, and it is also the infrastructure for city informatization. In the digital city phase, the main content of the geospatial framework is presented as a basic geographic information database and a public geographic information platform. With the development of the Internet of Things and cloud computing technologies, the basic geographic information database has become a spatiotemporal information database, and the



TABLE 2: Industry sector proportions of digital economy in smart city construction.

| Industry sector  | Infrastructure of enterprises | Electronic business | Electronic commerce | Computer network |
|------------------|-------------------------------|---------------------|---------------------|------------------|
| Manufacturing    | 0.738                         | 0.379               | 0.584               | 0.129            |
| Transportation   | 0.569                         | 0.448               | 0.172               | 0.245            |
| Public utilities | 0.642                         | 0.285               | 0.481               | 0.176            |
| Construction     | 0.723                         | 0.316               | 0.313               | 0.441            |
| Wholesale        | 0.614                         | 0.408               | 0.236               | 0.213            |
| Retail           | 0.358                         | 0.215               | 0.141               | 0.277            |
| Tourism          | 0.657                         | 0.226               | 0.563               | 0.545            |
| Catering         | 0.365                         | 0.466               | 0.370               | 0.428            |
| Information      | 0.812                         | 0.354               | 0.421               | 0.397            |
| Communication    | 0.764                         | 0.265               | 0.304               | 0.143            |

public geographic information platform has become a spatiotemporal information cloud platform. In the smart city stage, the geospatial framework has spatiotemporal characteristics and has evolved into a spatiotemporal information framework. The core content includes a spatiotemporal information database and a spatiotemporal information cloud platform.

For information, it is mainly to use language, characters, numbers, symbols, and other media to form special symbols to represent the content and characteristics of objective things; for data, it is mainly the carrier of information. In building a smart city, different methods need to be used to collect information, organize information, and analyze information. There are many types of information, such as natural information, social information, humanistic information, geographic information, etc. The transformation of digital cities into smart city is an inevitable requirement for the development of digital cities and an important way for cities to realize informatization (Figure 4). In the construction of digital cities, a variety of information has been gathered to ensure the normal construction, planning, management, and operation of the city. When building a smart city, the results of digital cities should be fully used to apply the information and data of digital cities. In the construction of smart city, this will promote the construction of smart cities. The construction of a smart city is to better and more scientifically manage the city, so as to improve people's living standards and let people feel the results brought by the smart city. Therefore, when building a smart city, the management decision-making platform should be optimized and improved on the basis of a digital city, so as to promote the application of smart city in transportation, medical treatment, and public safety spatial geographic information [17].

The smart city will multiply the city economy to some extent. Firstly, the research and development, manufacturing, and application of the Internet of Things, the Internet, supercomputers, and data processing and computing technologies and these high-tech capital-intensive industries will become new engines of economic growth and industrial upgrading and will generate a wide range, large markets, smart industrial chain, and industrial clusters with long chains and multiple connections. Secondly, while promoting the upgrading of city operating systems, a smart city will also make the wealth creation activities based on city

operating systems more convenient and efficient. A smart city is a product of the combination of a virtual economy and a real economy, and it will bring profound changes to city development. The construction of a smart city, through the widespread application of smart technologies, can increase the contribution of knowledge, technology, and information resources to economic growth. It can not only save material resources, reduce resource and energy consumption, reduce environmental pollution, and improve resource allocation efficiency, but also promote the transformation of economic development from labor-intensive and capital-intensive to knowledge-intensive and technology-intensive and accelerate the optimization and adjustment of economic structure. As a result of transformation and upgrading, some intelligent industrial chain and industrial groups with a wide range, large market, long chain, and many connected have been born; economic development is more intelligent, which in turn enhances the city's innovation and competitiveness.

## 5. Discussion

*5.1. Multifactor Analysis of Economic Growth in the Context of Smart City Construction.* The foundation of a smart city is a digital city, which can innovate in a measurable and visualized management and operation form of the digital city and create a networked comprehensive city decision-making and city management platform in the city's data and information infrastructure. The wireless network and the data network constitute a wireless sensor network, which is data-centric and designed for a certain requirement. The wireless sensor network refers to the realization of real-time monitoring of each node based on the wireless network for a certain characteristic application and can collect the information of the detection object and the environment information in the network and then process it to pass it to the user in more detail and comprehensive information (Figure 5). The network must have the ability to autonomously organize and reconfigure; when a node in the layout fails due to changes in the environment, the network topology changes over time. Therefore, the network is required to have the function of maintaining dynamic routing. Failures can occur due to node failure. Intelligent transportation systems can improve the efficiency of transportation systems, realize intelligent, networked, information and integrated transportation systems, and intelligently collect traffic flow, information, noise, and traffic accidents. The

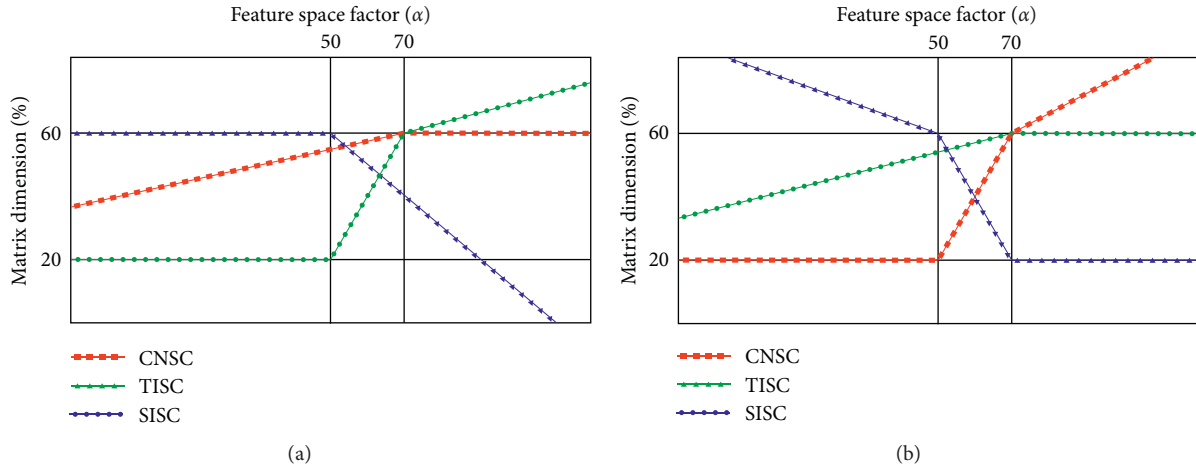


FIGURE 4: Relationship between matrix dimension and feature space factor of smart city's construction model based on digital economy. CNSC, network connection of smart city; SISC, service interaction of smart city; TISC, technology infrastructure of smart city.

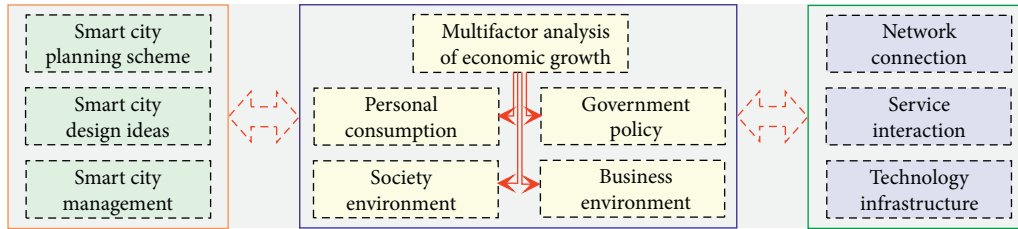


FIGURE 5: Multifactor analysis of economic growth in the context of smart city construction.

environments can communicate with each other to improve the safety and economy of the transportation system.

The intelligent applications will facilitate the management of the city and the life of citizens, while also collecting extensive user information, which will become the basis for planning, resource deployment, research, and analysis by government and other decision-making agencies. In the field of a smart city, data has a very important status, including geospatial data, industry data, census data, sensor monitoring data, etc., which are the basis of smart city construction [18]. The market demand is analyzed from infrastructure of enterprises, electronic business, electronic commerce, and computer network. The technological innovation environment of market demand must be first optimized and the public integrated process of the service platform then should be improved to guide the whole society to actively innovate and start businesses. The proposal of the integration of cities and industries is an inevitable requirement for responding to the transformation of industrial functions and the improvement of comprehensive city functions. It also reflects the trend of city planning from functionalism to humanism. From the focus on functional zoning and industrial structure, the focus is on integration and development and also is on people's initiative and the transformation of innovation and development. The smart cities emphasize the high degree of integration of industrialization and informationization and rely on the development of highly developed information infrastructure to create conditions for the extensive interconnection and

interconnection of city big data. The digital economy will build a smart park into a comprehensive high-tech park that integrates technology research and development, public platforms, demonstrations, system integration, and consulting services. With the help of geographic information technology, the citizen information terminal integrates city management information and establishes a precise, efficient, comprehensive, and visual city management model.

The smart city construction is a systematic project that requires a large amount of funds to support it. Therefore, it is necessary to broaden the financing channels, especially to promote the participation of social capital. The government should accurately convey the value information of smart city construction and, for social capital investors, then accelerate the financing of smart city construction through energy-saving bonds, green bonds, etc. (Figure 6). This is a feasible strategy to promote the development and development of smart city. Information and communication are two key technologies in the construction of smart city, so they cannot be measured according to a specific technical term. Instead, they should be judged based on value functions. For example, in the process of clarifying water standards, network speed and coverage of the scope, penetration rate, timeline, etc. are clearly defined so as to better meet the needs of functional parameters in the construction of smart city under clear technical standards to achieve convergence with the development of future cities [9]. In addition, the smart city construction is a systematic project, so during the construction process, a management system that can be used

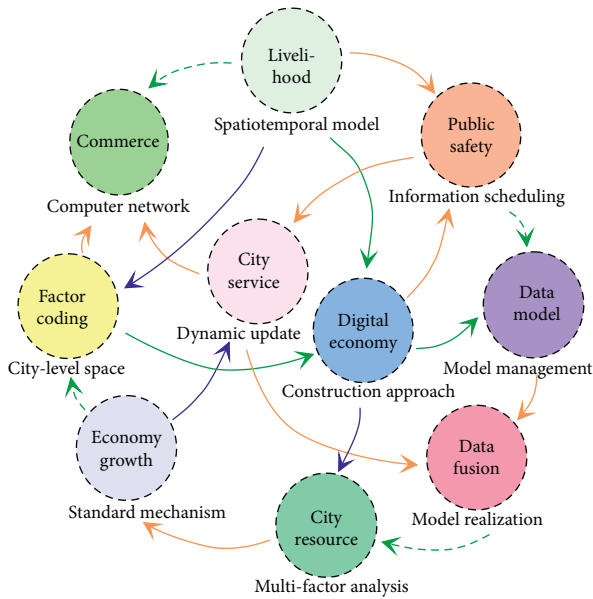


FIGURE 6: Realization of people-oriented social management model based on digital economy.

by decision-makers to judge success or failure should be established first, so as to continuously evaluate the plan progress, and it should also provide a certain early warning function, so as to ensure the steady progress of smart city construction.

### 5.2. Standardization Construction Mechanism of Smart City.

The competition for smart city development includes competition for industries, markets, talents, and investment and development opportunities; and the development has the corresponding effect, and resources will quickly gather to the core. When a place is successful, it will close the possibility of neighboring cities to follow the same road and an area cannot form two centers with the same characteristics. The success of regional development strategies is often not entirely determined by their own wishes and the development strategies of surrounding cities will affect the implementation of established strategies. The behavior of other regions is unpredictable and these uncertainties make static city planning extremely risky. At present, the emergence of industrial development plans in cities will inevitably greatly reduce the success rate of the plans. Because city information industry development planning is the field with the highest isomorphism, localities generally lack ideas and innovations in planning software, service outsourcing, digital content industry, Internet industry, and Internet of Things industry. The generation of isomorphic planning is based solely on the results of digital analysis and the same trend data can only lead to the same recommendations. The information industry planning is a risk decision in the face of uncertainty (Figure 7). The engineering logic cannot give innovative power, which suggests that the regional industrial planning must be independent and take risks [19].

The information service providers in the middle of the industry chain mainly include various information technology companies and network operation providers. The information service providers include software and application development, information integration and transmission, large-scale data processing, and cloud computing. The network operation provider is mainly responsible for project investment, construction, and operation and then resells or leases it to the government to provide basic communication and broadband network operation functions. The operation and service providers are important links in the industrial chain and midstream smart city construction system integrators present a trend. They have begun to move from the field of smart applications to the top-level design of smart city and from small integration of some applications to large integration of smart city construction, trying to establish standards and seize the right to operate a smart city (Figure 8). The industrial chain for the construction of a smart city contains all the content for the development of a smart economy industry and becomes an operating platform to realize data monetization. The key to smart economy is the development and integration of the smart city construction industry chain and the key to the latter lies in the coordinated and innovative development of the industry. Through the penetration, crossover, and reorganization of information, technology, talents, funds, and other resources, the entire wisdom is reconstructed. The industrial chain of city construction and even the city system promotes the development of smart industries and forms a smart economy.

The smart cities are digitally networked, intelligent, interactive, and synergistic at the technical level and they are especially demanding for the level of interconnection and interoperability of various city systems, at the social, economic, cultural, and educational levels. This emphasizes people-centered sustainable innovation and focuses on the collaborative sharing of a knowledge-based society, which involves a wide range of aspects, which determines that the construction of a smart city urgently requires the comprehensive use of various information technologies and products, as well as the innovation of business operations and public service models [20]. Therefore, the best order function of standard formulation can be used to solve problems such as data fusion and sharing in the construction of smart city, and the use and reuse of standardized activities to build, innovate, and promote related models of smart city construction. The mechanism between the two relationships can be developed from the following aspects and the construction of a smart city is a new thing, and it is necessary to constantly explore and accumulate experience in the construction. Among them, the promotion of replication and demonstration pilots is an effective way. In this process, standardization can transform the valuable and fully available results of the pilot into standard specifications in a timely manner, form a standardized working model for smart city construction, and promote the demonstration pilots to be carried out in a wider and deeper field.

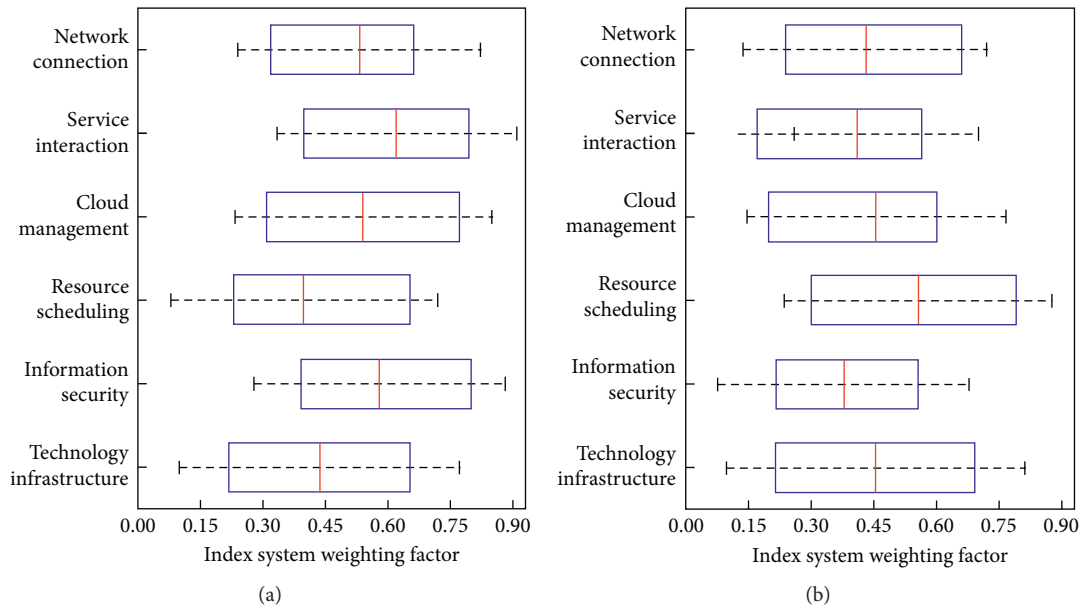


FIGURE 7: Index systems weighting factors of different indicators in the resource scheduling (a) and strategic management (b) of smart city based on digital economy.

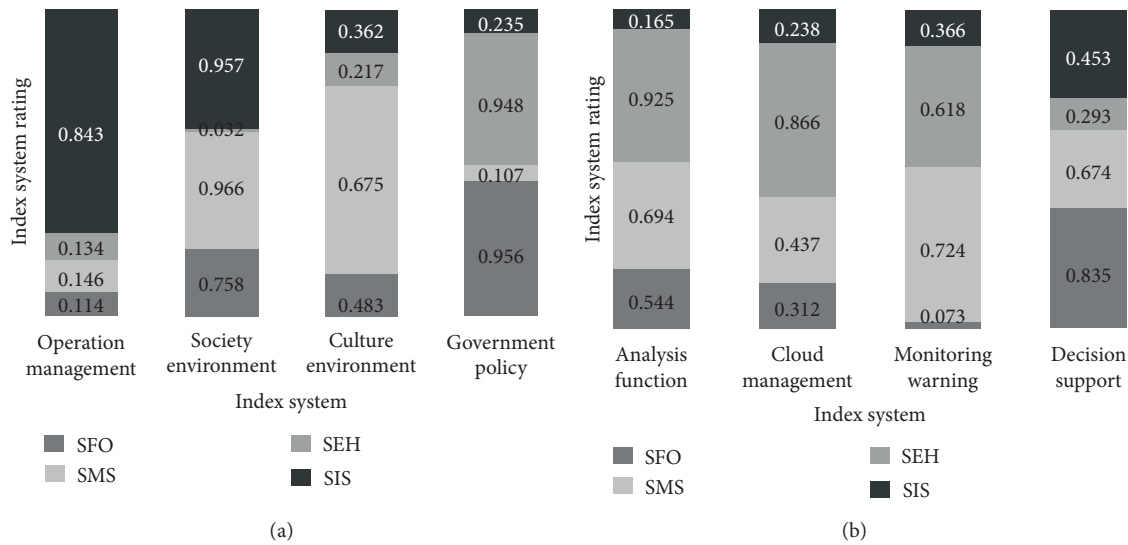


FIGURE 8: Index system rating of the resource scheduling (a) and strategic management (b) of smart city based on digital economy. SFO, smart foundation operation; SMS, smart management services; SHE, smart economy and humanity; SIS, smart integrated support.

## 6. Conclusions

This paper formulated resource scheduling strategies for a smart city based on digital economy, explored the dynamic fusion, storage, and update of a smart city’s multisource heterogeneous data, conducted the information display and analysis of multilevel smart city, proposed strategic management approaches for a smart city based on digital economy, analyzed the integrated implementation model of shared resource scheduling and people-oriented social management, and discussed the economic growth factors and standardization mechanism of a smart city under the background of digital economy. In the digital economy,

digital networks and communication technology facilities provide a global platform for individuals and organizations around the world to achieve mutual communication, communication, and cooperation. In the digital economy, transactions usually have the characteristics of digitization, virtualization, concealment, and electronic payment, which make it difficult to judge the consumption of trading venues and provide goods and services. The focus of smart city construction is to give play to the role of the market in allocating resources, to create various personalized smart applications that meet market needs through price, competition, and other factors, to foster a good state of emerging markets, and to promote new growth in the smart city

economy. The core of a smart city is a smarter way to change the way governments, communities, or companies and people interact with each other through a new generation of information technology to improve the clarity, efficiency, flexibility, and response speed of interactions and the characteristics are perception, interconnection, and intelligence. The results of this study provide a reference for further researches on the resource scheduling and strategic management of a smart city under the background of digital economy.

### Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Acknowledgments

This work was supported by the project of the National Social Science Fund (No. 17BGL030) and Youth Scientific Research and Innovation Fund Project of Henan University of Animal Husbandry and Economy (No. XKYCXJJ2017006).

### References

- [1] F. Li, A. Nucciarelli, S. Roden, and G. Graham, "How smart cities transform operations models: a new research agenda for operations management in the digital economy," *Production Planning & Control*, vol. 27, no. 6, pp. 514–528, 2016.
- [2] O. O. Komarevtseva, "Smart city technologies: new barriers to investment or a method for solving the economic problems of municipalities?" *R-economy*, vol. 3, no. 1, pp. 32–39, 2017.
- [3] A. Glasmeier and S. Christopherson, "Thinking about smart cities," *Cambridge Journal of Regions, Economy and Society*, vol. 8, no. 1, pp. 3–12, 2015.
- [4] T. Shelton, M. Zook, and A. Wiig, "The "actually existing smart city"" *Cambridge Journal of Regions, Economy and Society*, vol. 8, no. 1, pp. 13–25, 2015.
- [5] A. Wiig, "The empty rhetoric of the smart city: from digital inclusion to economic promotion in Philadelphia," *Urban Geography*, vol. 37, no. 4, pp. 535–553, 2016.
- [6] R. Kitchin, "Making sense of smart cities: addressing present shortcomings," *Cambridge Journal of Regions, Economy and Society*, vol. 8, no. 1, pp. 131–136, 2015.
- [7] R. Khatoun and S. Zeadally, "Cybersecurity and privacy solutions in smart cities," *Institute of Electrical and Electronics Engineers Communications Magazine*, vol. 55, no. 3, pp. 51–59, 2017.
- [8] G. Ferrara, "Smart city: a geographical perspective," *Romanian Review on Political Geography*, vol. 8, no. 2, pp. 43–48, 2016.
- [9] R. G. Hollands, "Critical interventions into the corporate smart city," *Cambridge Journal of Regions, Economy and Society*, vol. 8, no. 1, pp. 61–77, 2015.
- [10] L. Anthopoulos, "Smart utopia VS smart reality: learning by experience from 10 smart city cases," *Cities*, vol. 63, pp. 128–148, 2017.
- [11] V. Kupriyanovsky, A. Ishumuratov, D. Namiot, D. Yartsev, N. Utkin, and D. Nikolaev, "Digital economy and the Internet of things—negotiating data silo," *International Journal of Open Information Technologies*, vol. 4, no. 8, pp. 36–42, 2016.
- [12] J. Sun, J. Yan, and K. Z. Zhang, "Blockchain-based sharing services: what blockchain technology can contribute to smart cities," *Financial Innovation*, vol. 2, no. 1, pp. 1–9, 2016.
- [13] P. Cardullo and R. Kitchin, "Being a "citizen" in the smart city: up and down the scaffold of smart citizen participation in Dublin, Ireland," *GeoJournal*, vol. 84, no. 1, pp. 1–13, 2019.
- [14] C. C. Snow, D. D. Håkonsson, and B. Obel, "A smart city is a collaborative community," *California Management Review*, vol. 59, no. 1, pp. 92–108, 2016.
- [15] D. Namiot, V. Kupriyanovsky, A. Samorodov, O. Karasev, D. Zamolodchikov, and N. Fedorova, "Smart Cities and education in digital economy," *International Journal of Open Information Technologies*, vol. 5, no. 3, pp. 56–71, 2017.
- [16] V. Albino, U. Berardi, and R. M. Dangelico, "Smart cities: definitions, dimensions, performance, and initiatives," *Journal of Urban Technology*, vol. 22, no. 1, pp. 3–21, 2015.
- [17] S. P. Caird and S. H. Hallett, "Towards evaluation design for smart city development," *Journal of Urban Design*, vol. 24, no. 2, pp. 188–209, 2019.
- [18] K. Kourtit, P. Nijkamp, and J. Steenbruggen, "The significance of digital data systems for smart city policy," *Socio-Economic Planning Sciences*, vol. 58, pp. 13–21, 2017.
- [19] N. Komninos, C. Bratsas, C. Kakderi, and P. Tsarchopoulos, "Smart city ontologies: improving the effectiveness of smart city applications," *Journal of Smart Cities*, vol. 1, no. 1, pp. 31–46, 2019.
- [20] T. Yigitcanlar, M. Kamruzzaman, L. Buys et al., "Understanding "smart cities": intertwining development drivers with desired outcomes in a multidimensional framework," *Cities*, vol. 81, pp. 145–160, 2018.

## Corrigendum

# Corrigendum to “Explosion Resistance of Three-Dimensional Mesoscopic Model of Complex Closed-Cell Aluminum Foam Sandwich Structure Based on Random Generation Algorithm”

Zhen Wang, Wen Bin Gu , Xing Bo Xie, Qi Yuan, Yu Tian Chen, and Tao Jiang

Army Engineering University of PLA, Nanjing 210007, China

Correspondence should be addressed to Wen Bin Gu; 1120122090@bit.edu.cn

Received 28 October 2020; Accepted 28 October 2020; Published 21 November 2020

Copyright © 2020 Zhen Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the article titled “Explosion Resistance of Three-Dimensional Mesoscopic Model of Complex Closed-Cell Aluminum Foam Sandwich Structure Based on Random Generation Algorithm” [1], there was an error in

Figures 1(c) and 2 due to the incorrect version of the code being used. The authors confirm that this does not affect the results and conclusions of the article, and the corrected Figures 1(c) and 2 are as follows:

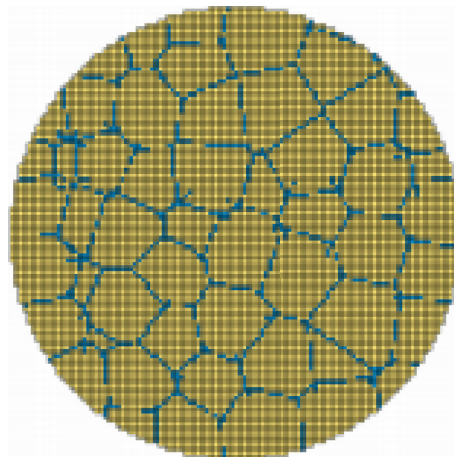


FIGURE 1: The real structure and numerical simulation structure of aluminum foam.

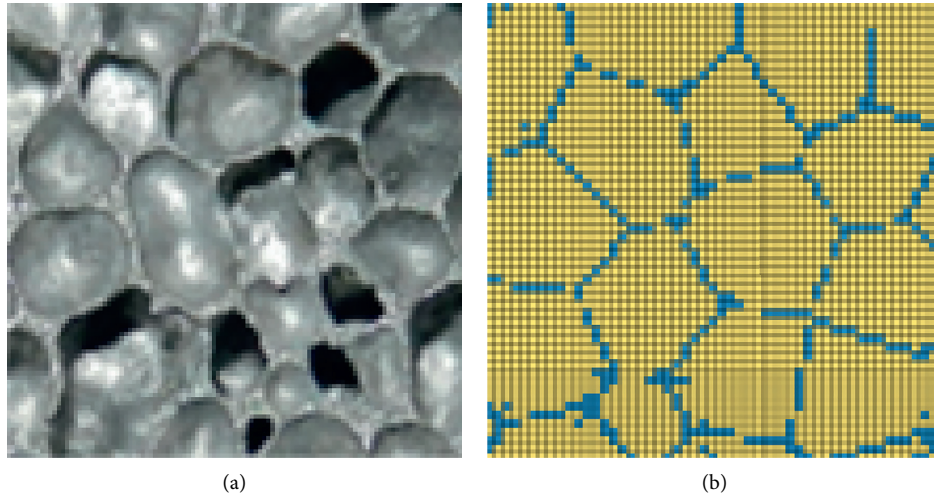


FIGURE 2: Comparison of different sections of the model with the real structure.

## References

- [1] Z. Wang, W. B. Gu, X. B. Xie, Q. Yuan, Y. T. Chen, and T. Jiang, "Explosion resistance of three-dimensional mesoscopic model of complex closed-cell aluminum foam sandwich structure based on random generation algorithm," *Complexity*, vol. 2020, Article ID 8390798, 16 pages, 2020.

## Research Article

# The Measurement of Chinese Sentence Semantic Complexity

Shuqin Zhu <sup>1,2</sup>, Jihua Song <sup>2</sup>, Weiming Peng <sup>2</sup>, Dongdong Guo,<sup>2</sup> and Jingbo Sun<sup>2</sup>

<sup>1</sup>Teacher's College of Beijing Union University, Beijing 100011, China

<sup>2</sup>School of Artificial Intelligence, Beijing Normal University, Beijing 100875, China

Correspondence should be addressed to Shuqin Zhu; [sftzhushuqin@bnu.edu.cn](mailto:sftzhushuqin@bnu.edu.cn), Jihua Song; [songjh@bnu.edu.cn](mailto:songjh@bnu.edu.cn), and Weiming Peng; [pengweiming@bnu.edu.cn](mailto:pengweiming@bnu.edu.cn)

Received 13 August 2020; Revised 19 September 2020; Accepted 6 October 2020; Published 20 November 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Shuqin Zhu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The complexity of language is usually reflected in the complexity of sentences. At present, the research of sentence complexity mainly focuses on the analysis of syntactic complexity. In this paper, from the perspective of Leech's theory of sentence semantic structure, the predication structure is taken as the semantic unit to explore the sentence semantic complexity. The predication structures are extracted based on the result of sentence-based syntactic analysis, and then the linear expression sequence of a sentence is converted into a semantic hierarchy based on predicate semantic frameworks; the universality of predicate semantic frameworks is obtained by using the spectral clustering algorithm; and the sentence semantic complexity depends on the universality of predicate semantic frameworks at various layers. The experimental results show that the measurement method of sentence semantic complexity based on predicate semantic frameworks is more effective by comparing with the method that only considers the semantic categories of words in the sentence.

## 1. Introduction

Language complexity refers to a property or quality of a phenomenon or entity in terms of (1) the number and the nature of the discrete components that the entity consists of and (2) the number and the nature of the relationships between the constituent components [1]. The complexity of language is embodied in vocabulary, pronunciation, grammar, and other subsystems. Among them, each plane subsystem (syntax, semantics, and pragmatics) within the grammar subsystem also has complexity [2]. This paper will focus on the semantic complexity, especially the measurement of sentence semantic complexity.

According to Leech's theory of sentence semantic structure, the predication structure is the main semantic unit of a sentence [3]. A predication structure can be divided into arguments and the predicate connecting arguments. Among them, the predicate is the main component of the predication structure, which determines the number and nature of arguments. Moreover, there are subordinate predication structures and degraded predication structures, and the difference between them lies in their different layers and

positions in sentences [4]. Yushu Hu pointed out the sentence semantics should not be sought from the lexical semantics in the sentence, but from the form or structure of the sentence. "Only by structural analysis can we summarize the common semantics from the same structures, and only by structural analysis can we find different semantics in different structures" [5].

According to the existing theory and analysis method of sentence semantic structure, this paper starts from the sentence structure and converts the linear expression sequence of the sentence into semantic hierarchy based on the results of sentence-based syntactic analysis. That is, the predication structure is used as the analysis unit. The predication structures of a sentence that need to be expressed preferentially are selected as the important parts, and the unimportant predication structures are selected as the additional components. The predication structures are arranged in layers according to the direct or indirect relationship between the various sentence components. Secondly, combined with the definition of words in HowNet [6], the arguments of the predication structures are further abstracted and generalized to obtain predicate semantic



frameworks (PSFs). In this way, the linear expression sequence of a sentence is converted into a semantic hierarchy, and the sentence semantic complexity is converted into the complexity of PSFs which are measured by the universality of PSFs. Spectral clustering is used to cluster the PSFs of a predicate, and the universality of PSFs in a large class is relatively higher. Finally, the sentence semantic universality depends on the universality of PSFs at each layer, and different weights are given to PSFs at different layers. The sentence semantic universality reflects the sentence semantic complexity. The sentences with high semantic universality are frequently used and the learning order is in the front. Sentences with low semantic universality make it difficult for learners to learn and understand [7, 8]. That is, the higher the sentence semantic universality is, the lower the sentence semantic complexity is.

The main innovations of this paper are as follows: one is to propose a measurement method of the universality of PSFs based on the predication structure, so as to obtain the universality of different PSFs of a predicate; the second is to propose an assessment method of sentence semantic universality based on PSFs, and the sentence semantic complexity is reflected by the sentence semantic universality.

## 2. Related works

At present, sentence complexity is mainly analyzed from structure and syntax. In [9], it is considered that two kinds of commonly used operations to complicate the content of clauses are parallel compound structure and nesting clause structure. Among them, the parallel compound structure takes the total number of commas and parallel conjunctions appearing in clauses as the quantitative estimation basis for difficulty, and the nesting clause structure takes the number of core verbs appearing in clauses as the estimation basis for difficulty. The mean of the difficulty estimation value of all clauses is taken as the difficulty estimation value of the sentence. In [10], a linear comprehensive evaluation model is used to calculate the complexity of Chinese structure. The indicators used in the model include the total number of clauses, the number of embedded or subordinate clauses in clauses, and the ratio of the word number to the clause number.

In addition, in the field of second language teaching, syntactic complexity is mainly used to measure the syntactic usage of learners' language output, which is an important indicator of learners' language level and language development trajectory. L2SCA is a syntactic complexity analysis tool for English second language, which covers 14 indicators including 5 dimensions of syntactic length, dependency, collocation, phrase complexity, and sentence overall complexity [11, 12]. Paper [13] also selects 14 measurement indicators from three categories and five subcategories for the syntactic complexity of Chinese as a second language, namely, the number of characters, words, syntactic components, phrases, clauses, consortiums, partial relations, complement structures, conjunctions, disjunctions, disposals, and passive, existential, and relative clauses in a basic unit. Papers [14–19] also study sentence complexity, and

researchers try to use various quantitative indicators to quantify sentence complexity.

Most of the existing researches on syntactic complexity focus on the analysis of sentence structure and formal features. Biber believes that simply considering sentence complexity from the perspective of structure does not really reflect its essence [20]. Ortega also believes that the semantic, function, and communicative value of sentence complexity should be analyzed and studied [21]. In addition, according to Bulté and Alex Housen, the complexity of language learning cognition consists of at least three parts: proposition complexity, discourse interaction complexity, and language complexity [1]. Among them, proposition refers to the semantics expressed in the text, not just the statement itself. The semantic structure of a proposition can be expressed as a “predication structure.” Proposition complexity is a relatively new concept, which has received far less attention than language complexity [22, 23].

Therefore, this paper attempts to analyze the sentence semantic complexity based on the basic proposition. In Section 3, the extraction of predication structures, the acquisition of PSFs, and the calculation method of the universality of PSFs are introduced. In Section 4, the calculation method of sentence semantic universality is introduced. The experimental results are introduced and analyzed in Section 5. Finally, the conclusion and limitations of this study are discussed in Section 6.

## 3. Universality of PSFs

The calculation method of the universality of PSFs is shown in Figure 1. Based on the results of sentence-based syntactic analysis, the predication structures are extracted layer by layer, and the PSFs are obtained by combining the definition of words in HowNet. All the PSFs of a predicate are clustered to get the universality of the PSFs. In addition, it is necessary to calculate the similarity of PSFs through lexical similarity and sememe similarity in order to cluster PSFs.

*3.1. Extraction of Predication Structures.* The extraction of predication structures is based on the result of syntactic analysis in the sentence-based treebank [24, 25]. The analysis and annotation of sentences in the sentence-based treebank are in the form of visual diagram, as shown in Figure 2. The horizontal line is the benchmark to observe the sentence layer. The subject, predicate, object, attribute, adverbial, complement, and other sentence components attached to the same horizontal line belong to the same layer. The subject, predicate, and object are located above the line, which are the “main components” of the sentence pattern; the attribute, adverbial, and complement are located below the line, which are the “additional components” of the sentence pattern; for the complex additional components, the syntactic analysis goes deep layer by layer. The annotation results are stored in XML form. The diagram and XML can be transformed in both directions.

Based on the results of sentence-based syntactic analysis, the long horizontal line with predicate component is taken

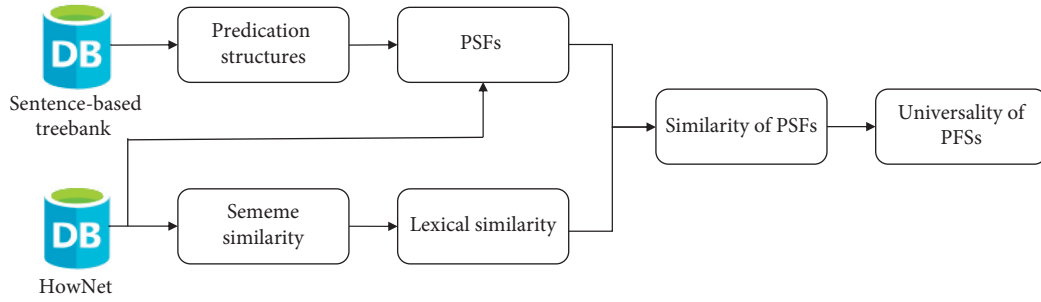


FIGURE 1: The calculation method of the universality of PSFs.

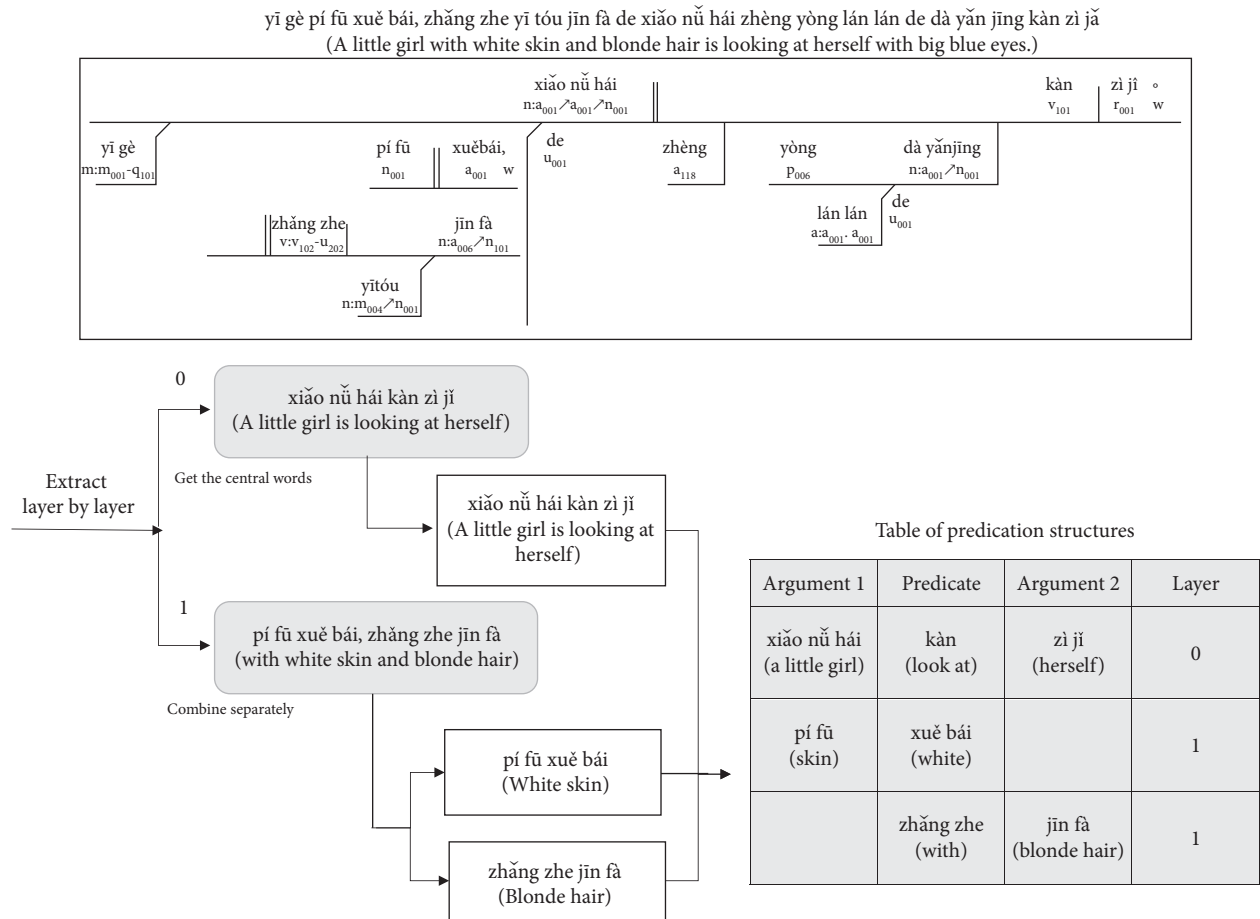


FIGURE 2: The extraction of predication structures.

as the baseline to extract the central word sequence directly related to the predicate. After the central word sequence of each layer is obtained, the predication structures are obtained by splitting and combining multiple predicates, and the process is shown in Figure 2.

It is possible that there are juxtaposed components in the subject or object. At this time, each component needs to be combined with core predicate separately. For example, in the sentence “yán sè, yàng zi dōu bǐ gāng cái kàn de qí páo hǎo (The color and style are better than those of the cheongsam I saw just now),” the subject includes juxtaposition, namely, “yán sè (color)” and “yàng zi (style).” The

predication structures of layer 0 are “yán sè hǎo (The color is good)” and “yàng zi hǎo (The style is good).”

The sentences with multiple predicates need to be split. Table 1 lists the split methods of the compound predicates, joint predicates, linked predicates, and pivotal sentence.

Considering the complexity of Chinese language, sentence components not only are acted by words, but also may contain a new predication structure, which is directly identified by the “VP.” For example, in the predication structure at layer 0 of the sentence “lì shǐ yǐ jīng zhèng míng tā zhǔ zhāng huáng quán shì cuò de (The history has proved that he is wrong in claiming imperial power),” “zhèng

TABLE 1: The split methods of multipredicates.

| Type                | Sentence  | Split results                             |
|---------------------|---|---|
| Compound predicates | wǐ shì fú nǐ le(I follow you!)                              | wǐ  shì<br>wǐ  fú  nǐ                     |
| Joint predicates    | mǔ qīn qiú shén bài fó(Mother prays for God and Buddha)     | mǔ qīn   qiú  shén<br>mǔ qīn    bài fó    |
| Linked predicates   | tā zhuā zhù wǐ de shǒu bù fàng (He held my hand)            | tā    zhuā zhù   shǒu<br>tā     bù   fàng |
| Pivotal sentence    | wǐ qǐng nǐ chī zhōng cān (I invite you to eat Chinese food) | wǐ    qǐng   nǐ<br>nǐ    chī   zhōng cān  |

míng(proof)” is the predicate and “lǐ shǐ(history)” and “VP” are arguments.

3.2. *Acquisition of PSFs.* Based on HowNet, the predicate structures are transformed into the PSFs. HowNet is a common sense knowledge base, which takes the concepts represented by Chinese and English words as the description object, and reveals the relationship between concepts and their attributes. HowNet defines a word as follows:

$$\textcircled{1} \text{wǐ:}\{\text{human}|\text{rén:PersonPro} = \{\text{1stPerson}|\text{wǐ}\}\}$$

$$\textcircled{2} \text{wǐ:}\{\text{specific}|\text{tè dìng:PersonPro} = \{\text{1stPerson}|\text{wǐ}\}\}$$

The first sememe in the definition of a word is the basic sememe, which points out the most basic meaning of the concept, such as “wǐ” referring to “human” or “specific.” The colon is followed by a detailed explanation of the basic sememe.

Combined with the semantic definition of words in HowNet [6], the PSFs can be obtained by abstracting and generalizing the arguments of predication structures, as shown in Table 2. Each word only takes the first basic sememe of each definition. Since it is impossible to know the exact semantics of each argument, if a word has multiple definitions in HowNet, all definitions in HowNet will be listed here for use in subsequent steps. If the word is not defined in HowNet, the word is used directly.

3.3. *Sememe Similarity.* Sememe similarity is the basis of calculating lexical similarity. Sememe similarity can be obtained by calculating sememe distance [26]. The most classical calculation method is as follows:

$$\text{sim}(s_1, s_2) = \frac{\alpha}{\text{dis}(s_1, s_2) + \alpha} \quad (1)$$

$\text{dis}(s_1, s_2)$  is the distance between  $s_1$  and  $s_2$  in the sememe tree. If  $s_1$  and  $s_2$  are in the same tree, the distance is the sum of the path lengths from  $s_1$  and  $s_2$  to their minimum common sememe. If  $s_1$  and  $s_2$  are not in the same tree, the distance will take a maximum of 20;  $\alpha$  is an adjustable parameter.

In the above calculation method, the weight of all paths is set to 1, but in HowNet, the difference between the top classes is large; the difference between the bottom classes is small. In view of this situation, [27] not only considers the depth of sememe tree, but also considers the regional density

TABLE 2: PSFs.

| Argument 1        | Predicate       | Argument 2  | Layer |
|-------------------|-----------------|---|-------|
| {human <br>rén}   | kàn(look at)    | {human rén}<br>{inanimate  wú shēng wù}<br>{self  jǐ} | 0     |
| {part bù<br>jiàn} | xuě bái(white)  |   | 1     |
|                   | zhǎng zhe(with) | {AppearanceValue <br>wài guān zhí}<br>{part bù jiàn}  | 1     |

of sememe tree. The calculation method of sememe similarity is as follows:

$$\text{sim}(s_1, s_2) = \frac{\alpha}{d + \alpha},$$

$$d = \delta \cdot \frac{\text{dis}(s_1, s_2)}{\text{con}(s_1) + \text{con}(s_2)},$$

$$\text{con}(s_1) = \gamma \text{deep}(s_1) + \eta \text{desity}(s_1) (\gamma < \eta \text{ and } \gamma + \eta = 1),$$

$$\text{desity}(s_1) = \frac{nc(s_1)}{\beta},$$

(2)

where  $\text{dis}(s_1, s_2)$  is the distance between  $s_1$  and  $s_2$  in the sememe tree.  $\text{deep}(s_1)$  is the depth of  $s_1$  in the sememe tree, that is, the path length from the root node to the sememe  $s_1$ .  $nc(s_1)$  is the sibling node number of  $s_1$ . The parameters are set as follows:  $\alpha = 1.6, \beta = 50, \gamma = 0.3, \eta = 0.7, \delta = 3$ .

3.4. *Similarity of PSFs.* There may be  $n$  parts (arguments) in a PSF. For two different semantic frameworks of a predicate ( $F_1$  and  $F_2$ ), if  $n$  is different, the possibility of similarity is small, and the similarity of the two PSFs is taken as 0. If  $n$  is the same, each framework has  $\text{ar}_1, \text{ar}_2, \dots, \text{ar}_n$  parts (arguments), and  $\text{sim}(F_1, F_2)$  is determined by the similarity of each part.

$$\begin{aligned} \text{sim}(F_1, F_2) &= \alpha_{\text{ar}_1} * \text{Sim}_{\text{ar}_1}(W_{F_1,1}, W_{F_2,1}) + \alpha_{\text{ar}_2} \\ &* \text{Sim}_{\text{ar}_2}(W_{F_1,2}, W_{F_2,2}) + \dots + \alpha_{\text{ar}_n} \\ &* \text{Sim}_{\text{ar}_n}(W_{F_1,n}, W_{F_2,n}). \end{aligned} \quad (3)$$

$\alpha_{ar1}, \alpha_{ar2}, \dots, \alpha_{arn}$  are the adjustable parameters, namely, the weight of each part, and  $\alpha_{ar1} + \alpha_{ar2} + \dots + \alpha_{arn} = 1$ . If  $W_{F_1,k}$  has  $m$  definitions in HowNet:  $S_{11}, S_{12}, \dots, S_{1m}$  and  $W_{F_2,k}$  has  $l$  definitions in HowNet:  $S_{21}, S_{22}, \dots, S_{2l}$ ,  $\text{sim}(W_{F_1,k}, W_{F_2,k})$  is the maximum value of similarity between definitions:

$$\text{sim}(W_{F_1,k}, W_{F_2,k}) = \max_{i=1\dots m, j=1\dots l} (\text{sim}(S_{1i}, S_{2j})). \quad (4)$$

For each part of a PSF, the first basic sememe of each definition is obtained from HowNet, so the similarity between definitions is the similarity between sememes.

The subject is the person or thing to be described in a sentence. It is the statement object of the predicate. The predicate and the object are generally combined to describe the subject. In view of the closer relationship between the predicate and the object, the parameters are set as follows:

- predicate + object + object(VOO) structure:  $\alpha_{ar1} = 0.5$ ,  
 $\alpha_{ar2} = 0.5$   
 subject + predicate + object(SVO) structure:  $\alpha_{ar1} = 0.2$ ,  
 $\alpha_{ar2} = 0.8$   
 subject + predicate + object + object(SVOO) structure:  
 $\alpha_{ar1} = 0.2, \alpha_{ar2} = 0.4, \alpha_{ar3} = 0.4$

**3.5. Clustering of PSFs.** The similarity matrix of PSFs is obtained by calculating the similarity between the semantic frameworks of each predicate. The method of spectral clustering is used to cluster the semantic frameworks of each predicate, and PSFs in large classes have a high universality.

Spectral clustering is a kind of clustering method based on graph theory [28–30]. All data vertices  $V = \{v_1, v_2, \dots, v_n\}$  form undirected weighted graph  $G(V, E)$ . Vertices can be connected by edges, and the weight  $w_{ij}$  on each edge represents the relationship between  $v_i$  and  $v_j$ . Because  $G$  is an undirected graph, the weight on the edges is independent of the direction of the two points,  $w_{ij} = w_{ji}$ . The matrix composed of the weights between any two points is the adjacency matrix  $W$  of a graph. For any point  $v_i$  in a graph, its degree  $d_i$  is defined as the sum of the weights of all the edges connected with it, that is,  $d_i = \sum_{j=1}^n w_{ij}$ . The degree matrix can be expressed as  $D$ .  $D$  is a diagonal matrix whose value is the degree of each vertex.

Each semantic framework of each predicate can be regarded as a vertex in graph  $G$ . The relationship between the semantic frameworks of each predicate is represented by the adjacency matrix  $W$ , that is, the PSFs similarity matrix of a predicate. Clustering is to cut the graph  $G$  into  $k$  subgraphs, so that the sum of edge weights between different subgraphs is as low as possible, while the sum of edge weights within subgraphs is as high as possible, as shown in Figure 3. The number of vertices contained in each subgraph is the universality of this kind of PSF  $u_i$ .

## 4. Sentence Semantic Universality

According to Levy, there are two different ways to understand sentences: one is based on memory; the other is based on expectation. Because of the need to complete the timely

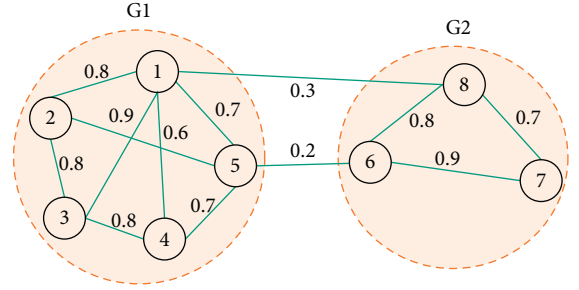


FIGURE 3: Clustering of PSFs.

storage, synthesis, and extraction of input information, it is difficult to understand based on memory [31]. The text that meets reading expectation is relatively easy to understand. For example, the following two sentences have the same number of words, but the premodifiers in the first sentence are juxtaposed, which meet reading expectation and are easy to understand. However, the second sentence is not easy to understand because of its multiple nesting of modifiers [10].

- (1) zài chù dǎngwěi de dà nǎo hóng 9 yuè, gēxīn 2 qiān jiàn, bǎozhèng bǎi mǐ jǐng, guóqīng bǎ lǐ xiàn de xíngdòng kǐuhào xià (25 words; under the slogan of the party committee's campaign to make a big splash in September, to innovate 2000 pieces, to ensure the 100-meter well and to present gifts on National Day).
- (2) duǎnduǎnde guānyú shìjiè shàng de zhǐngzhǐng de lǐshǐ de zōngjiào duìyú rénlèi de sǐwáng hòu de shēngmíng suǐ céngjīng qǔ guò de tàidù de xùshù (25 words; a short narrative about the attitudes of various historical religions in the world to human life after death).

Based on the above theory, the sentence semantic complexity can be divided into two parts: the complexity of the main PSFs and the complexity of the additional PSFs. Only by understanding the main PSFs can we grasp the central idea of the sentence. Only by clarifying the additional PSFs can we get a complete understanding of sentence semantics. Different weights are given to PSFs at different layers, and the semantic universality of a sentence ( $U_{sen}$ ) with  $n$  structures is the synthesis of the universality of PSFs ( $u_i$ ) in every layer.

$$U_{sen} = \sum_{i=0, \dots, n} \alpha_i u_i, \quad (5)$$

where  $\alpha_i$  is an adjustable parameter, that is, the importance of different PSF, which will be determined later by experiments.

## 5. Experiments and Discussion

**5.1. Experimental Data.** 244 volumes of international Chinese textbooks in the sentence-based treebank are selected to obtain the universality of PSFs, which includes 4,695 documents and 91,526 sentences (separated by · ? !).

*Boya Chinese* is selected to complete experiments of sentence semantic complexity. *Boya Chinese* contains 9

volumes of textbooks. The difficulty of these textbooks increases in turn, and they can be divided into primary, intermediate, and advanced. The details are shown in Table 3.

5.2. *Universality of PFSs*. Based on 91,526 sentences, 231,020 predication structures are extracted. 1,138 predicates with a frequency greater than 20 are clustered. The contour coefficient is used to measure the density and dispersion of the classes, so as to automatically select the number of clusters. The calculation method of the contour coefficient is as follows:

$$S = \frac{b - a}{\max(a, b)}. \quad (6)$$

For a predication structure,  $a$  is the average distance from other predication structures in the same category, and  $b$  is the average distance from the predication structures in the different categories closest to it. The overall contour coefficient is the average value of all the contour coefficients. The larger the contour coefficient is, the better the dispersion between classes is; the smaller the contour coefficient is, the worse the clustering effect is.

After clustering, the percentage of a kind of predication structure can be obtained. As shown in Table 4, in the predication structures of “*tí gāo* (improve)”, the first class of predication structures accounts for 6.7%, and the second class of predication structures accounts for 24.6%. Combined with the occurrence frequency of the predicate, the universality of each predication structure can be obtained. For predicates whose frequencies are less than or equal to 20, the universalities of their predication structures are set to 1.

5.3. *Sentence Semantic Universality*. This paper analyzes the sentence semantic universality of *Boya Chinese*. At the same time, the setting methods of adjustable parameter in the calculation formula of sentence semantic universality are compared in this experiment.

Method 1: the sentence universality takes the lowest universality of PSFs in the sentence.

$$\alpha_i = \begin{cases} 1, & \text{if } \min(u_i), \\ 0, & \text{other.} \end{cases} \quad (7)$$

Method 2: if there is only one layer of syntactic structure in a sentence, the weights of all the predication structures are the same; otherwise, the weight of predication structures at the backbone layer is 0.8, and the weight of predication structures at the additional layer is 0.2.

First of all, method 1 is used to set adjustable parameters. Table 5 shows the distribution of sentence semantic universality in textbooks at all levels. It can be seen intuitively that, with the increase of text difficulty, the proportion of sentences with low universality gradually increases, from 26.6% to 82.6%, and the proportion of sentences with high universality is declining sharply.

Method 2 is used to calculate the sentence semantic universality, and the distribution of sentence semantic universality in each textbook is shown in Table 6. From the results in the table, the distribution of sentences with semantic universality between 1 and 20 in the textbooks of Book 1 to Book 9 is not rising steadily. The distribution of sentences with semantic universality more than 1000 has not achieved the expected effect, and the distribution law is not obvious in all levels of textbooks.

In order to compare the difference of sentence semantic universality between the two methods on text difficulty, the relative entropy (KL distance) between adjacent level texts is calculated based on sentence semantic universality. KL distances are shown in Table 7. It can be seen that the sentence semantic universality calculated by Method 1 can better distinguish texts at all levels, and the KL distances between textbook texts at adjacent levels are larger, so Method 1 is used to obtain sentence semantic universality. The effect of Method 2 is not as expected. This may be because the split of the sentence is too detailed when obtaining the predication structures, resulting in the frequencies of synthetic predicates being higher, which affects the calculation of sentence semantic universality. For example, the sentence “*wǐ néng qù yóu yǐng*(I can go swimming)” is divided into “*wǐ néng*(I can),” “*wǐ qù*(I go),” and “*wǐ yóu yǐng*(I swim).” In this case, the frequencies of predicates such as “*néng*(can)” and “*qù*(go)” have increased a lot.

#### 5.4. Comparative Experiment

5.4.1. *Baseline*. From the above experiments, it can be seen that when sentence semantic universality is used to represent sentence semantic complexity, sentence semantic complexity has obvious distribution law in all levels of texts (Method 1). The method in this paper closely connects structure and semantic, extracts the predication structures layer by layer based on the results of syntactic analysis, and synthesizes the complexity of the predication structures at all levels of a sentence.

In order to further verify the effectiveness of this method, the following method does not consider sentence structure and only measures the sentence semantic complexity from the diversity of lexical semantics. The calculation method is given as an example below [32].

If there is a dialogue below:

- (A) *wǐ de bà bà yán jiū de shì shù xué, nǐ de bà bà ne?*  
(My dad studies mathematics, what about your dad?).
- (B) *wǐ de bà bà shì shū fǎ.* (My father studies calligraphy).

So, although the structure of the following two sentences is the same, it is clear that the first sentence is easier to understand than the second sentence, because the semantics of “*bà bà*(daddy)” and “*jūn rén*(military)” are the same [32].

- (1) *wǐ de bà bà shì jūn rén* (My father is a soldier).
- (2) *wǐ de bà bà shì shū fǎ* (My father studies calligraphy).

TABLE 3: Boya Chinese.

| Title                 | Character | Level        | Chapter | Sentence | Average of character in a sentence | Average of word number in a sentence |
|-----------------------|-----------|--------------|---------|----------|------------------------------------|--------------------------------------|
| <i>Boya Chinese 1</i> | 5876      | Primary      | 66      | 1075     | 12.244                             | 7.46                                 |
| <i>Boya Chinese 2</i> | 7259      |              |         |          |                                    |                                      |
| <i>Boya Chinese 3</i> | 11560     | Intermediate | 67      | 2549     | 24.774                             | 14.519                               |
| <i>Boya Chinese 4</i> | 17390     |              |         |          |                                    |                                      |
| <i>Boya Chinese 5</i> | 14475     |              |         |          |                                    |                                      |
| <i>Boya Chinese 6</i> | 28344     |              |         |          |                                    |                                      |
| <i>Boya Chinese 7</i> | 26421     | Advanced     | 48      | 2942     | 27.002                             | 15.035                               |
| <i>Boya Chinese 8</i> | 27955     |              |         |          |                                    |                                      |
| <i>Boya Chinese 9</i> | 34942     |              |         |          |                                    |                                      |

TABLE 4: Predication structures.

| Class | Argument 1                            | Predicate          | Argument 2                           | %                         |
|-------|---------------------------------------|--------------------|--------------------------------------|---------------------------|
| 1     | tā(it)                                | tí gāo(improve)    | lì yòng lǜ(utilization rate)         | 6.7                       |
| 1     |                                       | tí gāo(improve)    | dān chǎn(per unit yield)             | 6.7                       |
| 1     |                                       | tí gāo(improve)    | chǎn liàng(yield)                    | 6.7                       |
| 1     |                                       | tí gāo(improve)    | jǐng tǐ(alert)                       | 6.7                       |
| 1     |                                       | jié jìng(shortcut) | tí gāo(improve)                      | shōu shì lǜ(viewing rate) |
| 2     | dī shōu rù jiē céng(low income class) | tí gāo(improve)    | shēng huó shuǐ píng(living standard) | 24.6                      |
| 2     |                                       | tí gāo(improve)    | shè huì dì wèi(social position)      | 24.6                      |
| 2     |                                       | tí gāo(improve)    | rén kǒu sù zhì(population quality)   | 24.6                      |
| 2     |                                       | tí gāo(improve)    | mǎn yì dù(satisfaction)              | 24.6                      |
| 2     |                                       | tí gāo(improve)    | néng lì(ability)                     | 24.6                      |
| 2     |                                       | tí gāo(improve)    | shēng huó zhì liàng(quality of life) | 24.6                      |
| 2     |                                       | tí gāo(improve)    | sù zhì(quality)                      | 24.6                      |

TABLE 5: Distribution of sentence semantic universality in each textbook (Method 1).

| Textbook | 1–20  | 21–100 | 101–200 | 201–300 | 301–400 | 401–1000 | >1000 |
|----------|-------|--------|---------|---------|---------|----------|-------|
| 1        | 26.60 | 29.40  | 8.20    | 5.60    | 1.80    | 7.80     | 20.60 |
| 2        | 52.00 | 25.26  | 7.37    | 5.26    | 0.42    | 2.74     | 6.95  |
| 3        | 60.89 | 22.67  | 5.56    | 4.00    | 0.44    | 2.00     | 4.44  |
| 4        | 69.01 | 20.13  | 2.08    | 2.40    | 0.48    | 2.40     | 3.51  |
| 5        | 72.77 | 14.19  | 2.97    | 1.37    | 0.23    | 2.75     | 5.72  |
| 6        | 80.02 | 10.10  | 1.72    | 1.29    | 0.43    | 1.72     | 4.73  |
| 7        | 76.75 | 10.72  | 2.30    | 1.70    | 0.40    | 1.20     | 6.91  |
| 8        | 78.68 | 10.98  | 2.00    | 1.27    | 0.18    | 1.91     | 4.99  |
| 9        | 82.60 | 10.18  | 1.13    | 0.14    | 0.42    | 1.41     | 4.10  |

TABLE 6: Distribution of sentence semantic universality in each textbook (Method 2).

| Textbook | 1–20 (%) | 21–100 (%) | 101–200 (%) | 201–300 (%) | 301–400 (%) | 401–1000 (%) | >1000 (%) |
|----------|----------|------------|-------------|-------------|-------------|--------------|-----------|
| 1        | 10.20    | 19.20      | 11.60       | 8.40        | 3.00        | 19.00        | 28.60     |
| 2        | 13.26    | 21.47      | 13.89       | 9.05        | 5.26        | 19.37        | 17.68     |
| 3        | 12.89    | 22.00      | 11.56       | 7.78        | 6.44        | 24.67        | 14.67     |
| 4        | 12.46    | 23.48      | 9.27        | 7.67        | 7.19        | 22.68        | 17.25     |
| 5        | 14.87    | 17.62      | 10.30       | 5.72        | 5.03        | 23.34        | 23.11     |
| 6        | 22.66    | 17.83      | 8.92        | 5.69        | 4.73        | 20.62        | 19.55     |
| 7        | 23.85    | 18.84      | 7.72        | 6.51        | 4.61        | 19.54        | 18.94     |
| 8        | 19.51    | 16.88      | 9.98        | 5.54        | 4.26        | 23.96        | 19.87     |
| 9        | 24.05    | 18.39      | 6.79        | 4.81        | 5.23        | 18.53        | 22.21     |

TABLE 7: KL distances between textbooks of adjacent level.

| Text                    | Method 1 | Method 2 |
|-------------------------|----------|----------|
| Primary & intermediate  | 0.10832  | 0.011571 |
| Intermediate & advanced | 0.00822  | 0.005318 |

The semantics of each word in the sentences obtained from HowNet are as follows (because the semantic classification dictionary in [32] cannot be obtained, we count the number of semantic categories in the sentence based on HowNet):

- ①wǐ:human| rén
- ②wǐ:specific| tè dìng
- ③de:FuncWord| gōng néng cí
- ④bà bà: human| rén
- ⑤shì:be| shì
- ⑥shì:exist| cún zài
- ⑦shì:expression| cí yǔ
- ⑧shì:specific| tè dìng
- ⑨jūn rén: human| rén
- ⑩shū fǎ: method| fāng fǎ

Only the number of semantic categories is considered, and the occurrence number of semantic categories is not counted. The number of semantic categories in the first sentence (wǐ de bà bà shì jūn rén) is 6 (①②③⑤⑥⑦). The number of semantic categories in the second sentence (wǐ de bà bà shì shū fǎ) is 7 (①②③⑤⑥⑦⑩). In order to offset the influence of sentence length, the sentence semantic complexity = the number of semantic categories in the sentence / the number of words in the sentence [32]. The semantic complexity of the first sentence =  $6/5 = 1.2$ , and the semantic complexity of the second sentence =  $7/5 = 1.4$ . It can be seen that the second sentence has a higher complexity and is more difficult to understand.

**5.4.2. Results.** The summary of semantic complexity of sentences in *Boya Chinese* textbooks is shown in Table 8. The sentence complexity metrics obtained by the method in [32] and the method proposed in this paper are different. Using the method in [32], the representation of the sentence semantic complexity is ratio, the minimum is 0.5, the maximum is 12, and the median is 2.42. The representation of the sentence semantic complexity in this paper is frequency, with a median of 7.45.

In order to compare the two methods, the mapping functions of sentence semantic complexity are constructed firstly, and the sentence semantic complexity is divided into 1–6. The larger the value is, the more difficult the sentence is. After statistics and analysis of the distribution of sentence semantic complexity, the constructed mapping functions are shown in Table 9 (it should be noted that, after the analysis of the sentences in texts, it is found that the diversity of lexical semantics is less in the sentences of the more difficult texts, so monotonic decreasing function is also constructed).

The two methods are used to analyze the sentences in the textbooks (*Boya Chinese*) and calculate the average, standard deviation, and confidence interval of the sentence semantic complexity of each level of text (assuming that the distribution of sentence difficulty in each level of text follows Gaussian distribution, a 95% confidence interval is constructed). The results are shown in Table 10. It can be seen that as the difficulty of the text increases, the average of the sentence semantic complexity obtained by the two methods increases, but relatively speaking, the sentence semantic complexity obtained by the method proposed in this paper is better distinguished in all levels of text.

TABLE 8: Sentence semantic complexity in *Boya Chinese*.

|            | Min  | 1/4 quantile | Median | 3/4 quantile | Max     |
|------------|------|--------------|--------|--------------|---------|
| Paper [32] | 0.50 | 1.95         | 2.42   | 3.00         | 12.00   |
| This paper | 1.00 | 2.09         | 7.45   | 29.73        | 5285.50 |

TABLE 9: Mapping functions.

| Paper      | <b>a</b> | [0,1.5)  | [1.5,2)  | [2,2.5)   | [2.5,3)   | [3,3.5)    | [3.5,∞) |
|------------|----------|----------|----------|-----------|-----------|------------|---------|
| [32]       | $f_1(a)$ | 6        | 5        | 4         | 3         | 2          | 1       |
| This paper | <b>a</b> | $\leq 1$ | $\leq 5$ | $\leq 10$ | $\leq 20$ | $\leq 100$ | $>100$  |
|            | $f_2(a)$ | 6        | 5        | 4         | 3         | 2          | 1       |

TABLE 10: Comparison of sentence semantic complexity.

| Method     | Level of text | Average     | Standard deviation | Confidence interval |
|------------|---------------|-------------|--------------------|---------------------|
| Paper [32] | Primary       | 3.10        | 1.54               | [3.01,3.20]         |
|            | Intermediate  | 3.38        | 1.45               | [3.32,3.43]         |
|            | Advanced      | 3.57        | 1.47               | [3.51,3.62]         |
| This paper | Primary       | <b>2.43</b> | 1.43               | [2.34,2.52]         |
|            | Intermediate  | <b>3.76</b> | 1.63               | [3.70,3.83]         |
|            | Advanced      | <b>4.10</b> | 1.62               | [4.04,4.16]         |

TABLE 11: Correlation analysis of sentence semantic complexity and the text level.

| Parameters                      | Paper [32] | This paper |
|---------------------------------|------------|------------|
| Pearson correlation coefficient | 0.11       | 0.31       |
| T                               | 26.03      | 43.81      |
| Critical value (99%)            | 2.33       | 2.33       |

Due to the lack of Chinese sentence complexity tagging corpus, Pearson correlation coefficient is used to analyze the correlation between sentence semantic complexity and the text level. The results are shown in Table 11. The correlation coefficient of the method proposed in this paper is 0.31, which is significantly improved compared with the method of [32]. By constructing  $T$  to analyze the significance of correlation coefficient,  $T$  is not within the critical value ( $-2.33 < T < 2.33$ ), which indicates that there is a significant positive correlation between sentence semantics complexity and the text level at 99% confidence level.

The effect of measurement method based on predicate semantic frameworks is better than that only considering the number of semantic categories in sentences. The reason should be that the measurement method based on PSFs combines structure and semantics and takes predication structure as semantic unit, which not only measures the semantic collocation relationship and quantity between sentence elements from a horizontal perspective, but also examines the hierarchical system and the primary secondary relationship from a vertical perspective. It is a comprehensive analysis of the number and nature of elements in a language system, as well as the number of connections between these different elements.

## 6. Conclusion

Based on the results of sentence-based syntactic analysis, this paper extracts the predication structures and converts the predication structures into PSFs. The spectral clustering method is used to cluster the semantic frameworks of each predicate to obtain their universality. Then according to the number and importance of PSFs at different layers of the sentence, the sentence semantic universality is obtained. Experiments show that the sentence semantic universality can well reflect the sentence semantic complexity. Furthermore, the method is compared with the method that only considers the semantic categories of words in the sentence. Experimental results show that the proposed method in this paper can effectively measure the sentence semantic complexity.

In this paper, the universality of PSFs is only considered from the collocation universality of subject, object, and predicate, ignoring the relationship between adverbial, complement, and predicate. However, adverbial is the grammatical component that modifies the predicate, and complement is the component that complements and explains the predicate. They are closely related to the predicate. In addition, a predication structure is a reflection of the basic propositional semantic of the sentence. In addition to the basic propositional semantic, the sentence semantics also contain the superpropositional semantics, such as modal semantic, tense and aspect semantic, and degree semantic, which will be considered in the subsequent work.

## Data Availability

Sentence-based treebank and text corpus of international Chinese textbooks supporting this study have not been made available because the sentence-based treebank cannot be published until the relevant intellectual property protection application is completed. In addition, these textbooks belong to third party rights; the authors have no right to publish the data source.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grant nos. 61877004 and 62007004) and the Key Project of the National Social Science Foundation of China (Grant no. 18ZDA295).

## References

- [1] B. Bulté and A. Housen, *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*, John Benjamins, Amsterdam, The Netherlands, 2012.
- [2] R. Carston and D. Blakemore, "Introduction to coordination: syntax, semantics and pragmatics," *Lingua*, vol. 115, no. 4, pp. 353–358, 2005.
- [3] G. Leech, *Semantics: The Study of Meaning*, Penguin Books, Harmondsworth, UK, 2nd edition, 1981.
- [4] L. Si, "Research on sentence semantic generation," *Foreign Languages and Their Teaching*, vol. 184, no. 7, pp. 4–7, 2004.
- [5] Y. Hu and X. Fan, "Three planes of grammar research," *Language Teaching & Linguistic Studies*, vol. 2, pp. 4–21, 1993.
- [6] Z. Dong and Q. Dong, "HowNet-a hybrid language and knowledge resource," in *Proceedings of the 2003 International Conference on Natural Language Processing and Knowledge Engineering*, pp. 820–824, Beijing, China, October 2003.
- [7] E. Vyvyan, "Lexical concepts, cognitive models and meaning-construction," *Cognitive Linguistics*, vol. 17, no. 4, pp. 491–534, 2006.
- [8] V. Evans and J. Zinken, "Figurative Language in a Modern Theory of Meaning Construction: A Lexical Concepts and Cognitive Models Approach," in *Art, Body And Embodiment*, Cambridge Scholars Press, Cambridge, UK, 2007.
- [9] T. Mao, "Manual annotation approach to Chinese complex sentences by using bottom-up and top-down," *Journal of Chinese Computer Systems*, vol. 37, no. 4, pp. 716–721, 2016.
- [10] H. Qin and L. Kong, "The impact of translational Chinese on original language: a syntactic complexity perspective," *Journal of Foreign Languages*, vol. 41, no. 5, pp. 17–28, 2018.
- [11] X. Lu, "Automatic analysis of syntactic complexity in second language writing," *International Journal of Corpus Linguistics*, vol. 15, no. 4, pp. 474–496, 2010.
- [12] X. Lu and H. Ai, "Syntactic complexity in college-level English writing: differences among writers with diverse L1 backgrounds," *Journal of Second Language Writing*, vol. 29, no. SI, pp. 16–27, 2015.
- [13] Y. Wang, *A Study on the Measurement of Syntactic Complexity of Chinese as a Second Language*, Beijing Normal University, Beijing, China, 2015.
- [14] W. Jiang, "Measurements of development in L2 written production: the case of L2 Chinese," *Applied Linguistics*, vol. 34, no. 1, pp. 1–24, 2013.
- [15] J. E. Casal and J. J. Lee, "Syntactic complexity and writing quality in assessed first-year L2 writing," *Journal of Second Language Writing*, vol. 44, pp. 51–62, 2019.
- [16] B. R. Ambati, S. Reddy, and M. Steedman, "Assessing Relative Sentence Complexity Using an Incremental CCG Parser," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1051–1057, San Diego, CA, USA, June 2016.
- [17] F. Dell'Orletta, S. Montemagni, and G. Venturi, "Assessing document and sentence readability in less resourced languages and across textual genres," *International Journal of Applied Linguistics*, vol. 165, no. 2, pp. 163–193, 2015.
- [18] S. Jiang, *Research on Sentence Difficulty Measurement*, Xiamen University, Xiamen, China, 2009.
- [19] D. Yu, S. Wu, and C. Guo, "Assessing sentence difficulty in Chinese textbooks based on crowdsourcing," *Journal of Chinese Information Processing*, vol. 34, no. 2, pp. 17–26, 2020.
- [20] B. Douglas, G. Bethany, and P. Kornwipa, "Should we use characteristics of conversation to measure grammatical complexity in L2 writing development?" *Tesol Quarterly*, vol. 45, no. 1, pp. 5–35, 2012.
- [21] L. Ortega, "Syntactic complexity in L2 writing: progress and expansion," *Journal of Second Language Writing*, vol. 29, pp. 82–94, 2015.
- [22] H. Zaki and R. Ellis, "Learning Vocabulary through Interacting with Written Text," in *Learning a Second Language*



- through Interaction*, John Benjamins, Amsterdam, The Netherlands, 1999.
- [23] R. Ellis and G. Barkhuizen, *Analyzing Learner Language*, Oxford University Press, Oxford, UK, 2005.
  - [24] W. Peng, J. Song, Z. Sui et al., "Formal schema of diagrammatic Chinese syntactic analysis," in *Proceedings of the 16th Chinese Lexical Semantics Workshop*, pp. 701–710, Beijing, China, May 2015.
  - [25] S. Zhu, Y. Zhang, W. Peng et al., "Construction of the basic sentence-pattern instance database based on the international Chinese textbook treebank," in *Proceedings of 2016 International Conference on Asian Language Processing*, pp. 266–270, Tainan, Taiwan, November 2016.
  - [26] J. Xu, J. Liu, and Y. Zhang, "Word similarity computing based on hybrid hierarchical structure by HowNet," *Journal of Information Science and Engineering*, vol. 31, no. 6, pp. 2089–2101, 2015.
  - [27] X. Yuan, "Research on the calculation of semantic similarity of HowNet," *Journal of Liarning University (Natural Science Edition)*, vol. 38, no. 4, pp. 358–361, 2011.
  - [28] K. Li and Y. Liu, "A spectral clustering algorithm based on self-adaption," in *Proceedings of 6th International Conference On Machine Learning And Cybernetics*, pp. 3965–3968, Hong Kong, China, August 2007.
  - [29] H. Jia, S. Ding, X. Xu, and R. Nie, "The latest research progress on spectral clustering," *Neural Computing and Applications*, vol. 24, no. 7-8, pp. 1477–1486, 2014.
  - [30] C. Christina, V. Nicholas, and P. Ioannis, "Face clustering in videos based on spectral clustering techniques," in *Proceedings of 2011 First Asian Conference on Pattern Recognition*, pp. 130–134, Beijing, China, November 2011.
  - [31] L. Roger, F. Evelina, and G. Edward, "The syntactic complexity of Russian relative clauses," *Journal of Memory & Language*, vol. 69, no. 4, pp. 461–496, 2013.
  - [32] J. Zheng, "Lexical semantics and sentence difficulty measurement," in *Proceedings of Chinese Lexical Semantic Workshop*, pp. 261–265, Xiamen, China, April 2005.

## Research Article

# Three-Dimensional Finite Element Numerical Simulation and Analysis of Solid-State Processing of Metal Material

Guang Su<sup>1,2</sup> and Aimin Zhang <sup>1</sup>

<sup>1</sup>Department of Material Science and Engineering, Henan Institute of Technology, Xinxiang 453000, China

<sup>2</sup>Henan Engineering Research Center for Modification Technology of Metal Materials, Xinxiang 453003, China

Correspondence should be addressed to Aimin Zhang; zhangaimin@hait.edu.cn

Received 19 August 2020; Revised 27 October 2020; Accepted 2 November 2020; Published 16 November 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Guang Su and Aimin Zhang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Solid-state processing of metal material is a very complex physical and chemical process, which is coupled by a series of variations including heat transfer, momentum transfer, mass transfer, and phase change. Applying three-dimensional (3D) finite element numerical method to the simulation of solid-state processing can perform analysis of metal material's forging processes before production trial production, can obtain their relevant information such as material flow law, temperature field, and strain field under the minimum physical test conditions, thereby predicting metal material's forming defects and improving their forging quality. On the basis of summarizing and analyzing previous research works, this paper expounded the current status and significance of solid-state processing of metal materials, elaborated the development background, current status, and future challenges of 3D finite element numerical simulation, introduced the discrimination method and free surface solution method of numerical simulation calculation, conducted finite element model's geometric assumptions, material selection, element division, model establishment, parameter selection, and initial and boundary condition determination, and simulated and analyzed rheological casting, remelting heating, thixoforming, and rotary piercing processes of metal materials. The results show that the 3D finite element numerical method can not only simulate various processes of flow field, temperature field, stress field, and microstructure in solid-state processing but also can provide a reliable basis for effectively obtaining a reasonable description and finding a more optimized design plan for metal material processing in a short time, which plays an important role in understanding and analyzing solid metal forming process, controlling and optimizing process parameters, guiding and mastering rheological casting, and secondary heating and rotary piercing of metal materials.

## 1. Introduction

Solid-state processing of metal material is a very complex physical and chemical process, which is coupled by a series of variations including heat transfer, momentum transfer, mass transfer, and phase change. The simulation of temperature field and strain field in the solid-state processing of metal materials is to simulate the process of metal transforming from liquid to solid from the perspective of heat transfer, and the numerical solution of the thermal differential equation is obtained in the presence of phase change [1]. With the help of numerical simulation technology of temperature field and strain field, the formation mechanism of major casting defects, such as shrinkage

cavity, shrinkage porosity, thermal cracking, and macro-segregation, optimizes the design of the casting process, ensures the quality of castings, shortens the trial production cycle, and reduces production costs [2]. Solid-state processing of metal material refers to the deformability that a material can achieve without damage during plastic processing and is an important indicator of the material's bulk forming ability [3]. Material processing includes two aspects and the first one is stress state processing, which is related to the processing technology, mold and friction state, and can be controlled by changing the geometric size of the deformation area and the applied stress state; the other is internal workability, which reflects the decisive microstructure evolution under certain temperature, strain

rate, and strain conditions through the flow stress curve of the material and is very sensitive to the initial state of the material [4].

Applying three-dimensional (3D) finite element numerical method to the simulation of solid-state processing can perform analysis of metal material's forging processes before production trial production, can obtain their relevant information such as material flow law, temperature field, and strain field under the minimum physical test conditions, and thereby predicting metal material's forming defects and improving their forging quality [5]. The accurate simulation of these processes requires the constitutive equation of the liquid-solid mixture, and the contribution of the liquid to the deformation mechanism of mixture needs to be accurately described. Stress and strain field analysis is the basis for optimizing the rolling process. The key to accurate simulation is the establishment of solid material models and the optimization of simulation parameters, and the 3D simulation has further increased the difficulty [6]. Through numerical simulation, researchers can answer questions that cannot be answered in empirical design and understand the whole process of metal plastic forming, including the filling of materials at each stage of the metal forming process, the trend of material deformation, the stress, strain, and strain rate and forming inside the material load and velocity vector field [7]. This has great practical significance for the metal plastic forming process design, mold design, metal blank design, press selection, and forming quality control. Therefore, scholars have conducted a lot of research studies on the numerical simulation of metal material's solid-state processing and have obtained many research results [8].

On the basis of summarizing and analyzing previous research works, this paper expounded the current status and significance of solid-state processing of metal materials, elaborated the development background, current status, and future challenges of 3D finite element numerical simulation, introduced the discrimination method and free-surface solution method of numerical simulation calculation, conducted finite element model's geometric assumptions, material selection, element division, model establishment, parameter selection, and initial and boundary condition determination, and simulated and analyzed rheological casting, remelting heating, thixofforming, and rotary piercing processes of metal materials. The study results of this paper provide a reference for further research studies of 3D finite element numerical simulation and analysis of solid-state processing of metal material. The detailed chapters are arranged as follows. Section 2 introduces the discrimination method and free surface solution method of numerical simulation calculation. Section 3 conducts finite element model's geometric assumptions, material selection, element division, model establishment, parameter selection, and initial and boundary condition determination. Section 4 simulates the rheological casting and remelting heating processes of metal materials. Section 5 analyzes the thixofforming and rotary piercing processes of metal materials. Section 6 is conclusion.

## 2. Methods and Principles

*2.1. Discrimination Method of Numerical Simulation Calculation.* Rheological casting is actually a 3D heat conduction process, but the rheological casting used for solid-state processing of metal materials mainly prepares cylindrical billets, which is a symmetric problem. The differential equations for solving the thermal field problem of continuous casting semisolid cylindrical ingots can be written as follows:

$$\begin{aligned} & \frac{\partial(r_a T_x)}{\partial x} + \frac{\partial(r_b T_y)}{\partial y} + \frac{\partial(r_c T_z)}{\partial z} \\ & = f_s(x) \frac{\partial T_a}{\partial t} dL + f_s(y) \frac{\partial T_b}{\partial t} dL + f_s(z) \frac{\partial T_c}{\partial t} dL, \end{aligned} \quad (1)$$

where  $T_x$ ,  $T_y$ , and  $T_z$  are the thermal conductivity in the  $x$ ,  $y$ , and  $z$  directions, respectively;  $d$  is the material density;  $L$  is the latent heat of crystallization; and  $T_a$ ,  $T_b$ , and  $T_c$  are the specific heat;  $a$ ,  $b$ , and  $c$  are the axial displacement,  $f_s(x)$  The above differential equation can be solved by applying the finite element method or the difference method as the power consumption function.

According to the general viscosity model of early suspensions, the correlation between apparent viscosity and solid phase ratio and microstructure is

$$w = f_s \left[ 1 + \frac{-dL}{2(\lambda_k + 1/Cp)} \right], \quad (2)$$

where  $w$  is the apparent viscosity of the liquid alloy;  $\lambda_k$  is the critical solid phase rate, which mainly depends on the cooling rate and shear rate; and  $Cp$  is the specific surface area of the solid phase particles.

The central content of the theory of plasticity of metallic materials is the establishment of yield criterion. Since the complex physical characteristics of metal material deformation are difficult to describe with a very accurate mathematical model, the yield criterion based on the relationship between equivalent stress and equivalent strain increment is expressed as follows:

$$K = \frac{1}{\gamma} \sqrt{\frac{m}{2} \sigma + \frac{d^2 L^2}{w}} - P, \quad (3)$$

where  $K$  is the uniaxial yield stress;  $m$  is the stress component tensor;  $P$  is the hydrostatic pressure; and  $\sigma$  is the function of relative density, which can be obtained through uniaxial compression experiments and 3D compression experiments.

Previous studies have shown that the viscosity  $\eta$  and the shear rate of the alloy melt continuously cooled at a certain shear rate are [9]. The relationship of  $\gamma$  can usually be expressed as follows:

$$\eta = \left( K - \frac{f_s}{Cp} \right)^{-mr} \left\{ 1 + \left[ \frac{b(f_s)}{a} \right]^n \right\}^{n/k}, \quad (4)$$

where  $a$  and  $b$  are the critical solid fraction and instantaneous solid fraction related to the cooling rate and the shear

rate, respectively, and  $m$  and  $n$  are the consecutive times; the other parameters are empirical constants.

The thermal strain  $\delta_{ij}$  is related to the temperature and the thermal expansion coefficient of the material, and its expression is

$$\delta_{ij} = \begin{cases} \alpha_{ij}(T - T_\eta), & i = j, \\ \alpha_{ij}(T + T_\eta), & i \neq j, \end{cases} \quad (5)$$

where  $\alpha_{ij}$  is the thermal expansion coefficient of metal material;  $T$  is the temperature; and  $T_\eta$  is the reference temperature and one is the direct coupling unit. In addition to the displacement freedom of the ordinary unit, the unit also has the temperature freedom. After being affected by the local temperature, the temperature field in the unit is first calculated according to the heat conduction theory, and the temperature is taken as the temperature field. The external load is applied to the element and then combined with the element force and displacement constraints, and the internal force distribution of the element is obtained.

**2.2. Free Surface Solution Method.** The change of the free surface of liquid metal belongs to the category of interface dynamics, which is a difficult problem to deal with in the simulation of the metal flow process. In this method, the coordinate system is established on the fluid mass, and the grid moves with the mass, so the calculation method is simple and the determination of the free surface is more convenient. However, the flow and deformation of the mass will inevitably cause the grid to be distorted and distorted so that they are kinked with each other and even the grids may intersect, making the calculation impossible [10].

Since it is assumed that the solid material is equivalent to the porous material, the elastic deformation can be ignored. From the relationship between the strain rate tangent and the instantaneous stress, the rigid-constitutive equation of the solid material can be obtained as follows:

$$V_I = \gamma(I) \frac{\partial T}{\partial r}, \quad (6)$$

where  $V_I$  is the yield function and  $r$  is the material constant. The rheological behavior of solid metals is mainly affected by factors such as shear rate, solid phase volume fraction, solid phase particle morphology, and agglomeration degree, which can be expressed as follows:

$$h(x, y, z) = \eta(\gamma, f_s, m, c), \quad (7)$$

where  $\eta$  is the apparent viscosity,  $\gamma$  is the shear rate,  $m$  is the morphology parameter of solid particles, and  $c$  is the parameter of the degree of agglomeration of solid particles.

Therefore, when calculating the rolling force during the forming process, generally only the average total rolling force is calculated. The calculation formula is as follows:

$$B_e = h_a \sigma_b V_I (D + D_c) \sqrt{n \frac{\sigma_b}{V(D)}} e^{-n/L}, \quad (8)$$

where  $B_e$  is the average total rolling force,  $h_a$  is a parameter related to the taper of the core and uneven deformation,  $\sigma_b$  is a parameter that depends on the degree of hardening of the metal or alloy,  $D$  is the outer diameter of the rolled tube,  $D_c$  is the outer diameter of the tube,  $n$  is the amount of feed, and  $L$  is the length of the working part of the groove excluding the sizing section. If expressed in terms of heat capacity based on unit mass, the equation can be expressed as follows:

$$dW = mCp dT, \quad (9)$$

where  $m$  is the mass of the object and  $Cp$  is the heat capacity per unit mass, that is, the specific heat capacity of the object. The total power absorbed by the material per unit volume during thermal processing can be expressed by the sum of two complementary functions:

$$U = \int_0^v \sigma dv + \int_0^\sigma v d\sigma. \quad (10)$$

The  $U$  value in the formula represents the power consumption due to plastic deformation, most of which is converted into heat. The dissipation margin is the power consumption related to the change of the structure during the deformation of the material, so the change also represents the change of the microstructure.

### 3. Finite Element Model and Parameter Selection

**3.1. Basic Assumptions of the Geometric Model.** The assumptions for the solid-state processing of metal materials are as follows: the metal material is a compressible continuum; the elastic deformation of the material is ignored during the forming process of the metal material; the solid material is a non-Newtonian fluid. Thixoforming is to cool and solidify the semisolid billet obtained by special processes such as stirring, then cut the material according to the required size, then reheat it to the semisolid temperature, and then place it in the mold cavity for forming. The heating and conveying of semisolid metal blanks are very convenient, the forming process is easy to control, and it is convenient to realize automated production [11]. At present, most processes use thixoforming. After holding the alloy melt for a certain period of time near the liquids temperature, it is cast to obtain a semisolid slurry method suitable for thixoforming, and liquid casting alloy has a low melt temperature, almost no overheating, and a uniform temperature field. During the casting process, a large number of crystal nuclei are uniformly generated in the melt to form small, uniform, and semisolid slurry, and the method is simple, efficient, energy saving, and material saving. The alloy ingot with the structure is heated and sheared during the transportation process through a heating source and a special spiral propulsion system so that it has importance and is injected into the mold cavity to form injection molding which is essentially a kind of rheological technology.

**3.2. Material Selection and Unit Division.** This paper chooses TC11 alloy as the experimental metal material. TC11 alloy is a

kind of  $\alpha$ - $\beta$ -type heat-strength alloy with good comprehensive properties, which has excellent heat-strength properties below 500°C and high room temperature strength. The alloy is mainly used to manufacture parts such as compressor discs, blades, and drums of aero engines, as well as aircraft structural parts.

From the perspective of forging, the upper boss of the forging is not easy to be filled, so the upper boss is placed on the upper die. Although the shape of the bottom of the forging is complex, it is easy to fill. The size of the blank is obtained according to previous calculations and experiments and the initial heating temperature is 950°C. Considering the heat loss during the transfer process, the blank temperature is set to 930°C in the simulation. The size of the material is obtained according to previous calculations and experiments with its size of 400 mm × 120 mm. The blank is divided into 40,000 tetrahedral units (Figure 1).

**3.3. Establishment of Geometric Model.** Finite element software is used to simulate the forming process, and two kinds of blanks, cylindrical and stepped, are used for forming research, respectively. At the same time, in order to reduce the number of unit divisions and increase the calculation speed and time, only a certain proportion of the blank is taken as the simulation research object. Tetrahedral elements are used to discriminate the blank and redraw according to the grid distortion at any time during the calculation process to ensure the accuracy and convergence of the calculation. The points on the symmetry plane are constrained so that their normal velocity along the symmetry plane is zero (Figure 2). The deformation of the mold is not considered during the simulation, and its material property is a rigid body. In this paper, a simple and convenient parent element method is used to generate the grid of the initial blank. That is to say, the blank must be divided into simple sub-blocks and coordinate transformation, then transformed into a unit rectangle or a unit parallelepiped in the local coordinate system, and finally the grid is divided according to the specified grid division ratio in each direction.

During the deformation process, high temperature and large deformation are often generated in a local area to cause mesh distortion, which will reduce the accuracy of the solution or cause the mold to be embedded in the blank and penetrate, so adaptive mesh re-partitioning technology is required. As the deformation intensifies, the initial grid of the blank will be severely distorted, which may result in the inability to continue the calculation. Therefore, in order to overcome the computational difficulties and low accuracy caused by grid distortion, when the finite element grid is distorted to a certain extent, the calculation must be stopped, the grid suitable for the calculation must be redivided, and the modified boundary conditions must be determined and then continued to calculate. The automatic mesh redraw technology can correct the distorted or distorted mesh caused by excessive deformation. According to a certain set criterion, such as mold penetration, plastic strain, and incremental step criterion, it can automatically regenerate a

good-shaped mesh. The state variables in the original old grid are mapped to the newly divided grid to ensure subsequent calculations and improve calculation accuracy. Grid redrawing is generally divided into three steps of judgment of grid distortion, new grid generation, and data conversion between old and new grids [12].

### 3.4. Determination of Initial and Boundary Conditions.

The processing of the boundary conditions is the determination of the representative section of the tool and work-piece shape in the calculation, taking into account the geometric conditions, thermal effects, and other aspects. Such as work-piece cooling caused by tools or temperature rise caused by plastic deformation, as well as mechanical effects, including elastic flattening and bouncing of metal materials, are not considered. These simplifications are made to reduce the required calculation time, and the basic assumption makes the physical influence not play a major role when analyzing the influence of the equipment structure on the uneven flow of materials. These steps apply boundary condition constraints on geometric objects or finite element meshes, apply symmetric boundary conditions to corresponding symmetric nodes, and establish a set of two boundary nodes:  $Y=0$  node  $X$  displacement and  $Y$  and  $Z$  rotation is 0 and  $X=0$  node  $Y$  displacement and  $X$  and  $Z$  rotation is 0.

The metal material processing process mainly involves the heat transfer during the plastic deformation of the sheet surface. During the processing, the upper surface of the sheet material in contact with the air is a free surface, and the initial temperature of the sheet material and room temperature during the simulation is 24°C. The main heat exchange model is expressed as follows:

$$-k \frac{\partial T}{\partial x} = \lambda_c (T - T_0), \quad (11)$$

where  $k$  is the heat transfer coefficient between the metal material and the air and  $T$  is the absolute temperature. When calculating the temperature field, it is necessary to impose thermodynamic boundary conditions, including constant temperature boundary conditions and thermal convection boundary conditions. The constant temperature boundary condition simulates the surrounding external ambient temperature; the thermal convection boundary condition simulates the heat exchange on the contact surface between the structure and the air, assuming that the convective heat transfer coefficient is constant.

In the model, the speed of the contact surface with the cooling boot is 0, that is,  $v_x = 0$ ,  $v_y = 0$ , and  $v_z = 0$ . The speed of the contact surface with the work roll is the work roll speed:

$$\begin{aligned} v_{xi} &= v_0 \cos \alpha_i, \\ v_{yi} &= v_0 \sin \alpha_i, \end{aligned} \quad (12)$$

where  $v_0$  is the rotation speed of the work roll and  $i$  is the number of the node in the heat conduction matrix. The convective heat transfer coefficient on the entire contact

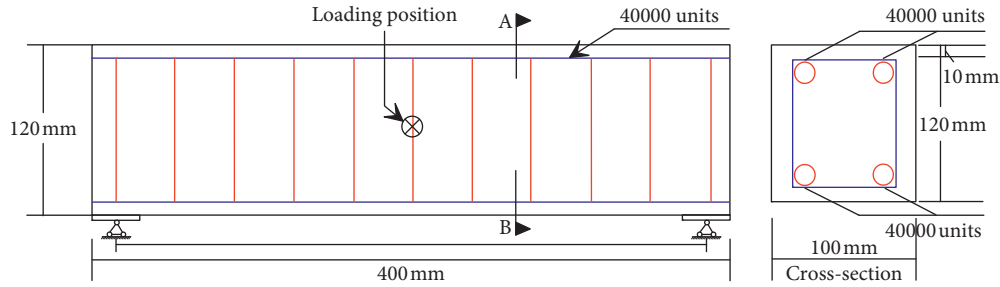


FIGURE 1: Metal material model and unit division for simulation and analysis of solid-state processing.

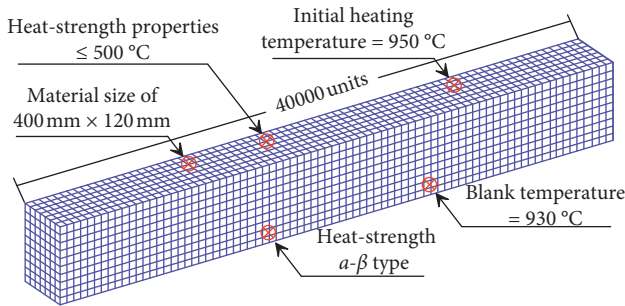


FIGURE 2: Establishment of the geometric model and its unit division of the metal material in solid-state processing.

surface changes with the temperature of the alloy and the mold and the contact state between the two, and it is accurately determined quite difficult. The metal material moves down at a constant speed, and the die adopts a floating structure, that is, during the pressing process of the metal material, the metal material forces the die to move downward together at the same speed. In the simulation process, it is assumed that the mold does not undergo plastic deformation and is a rigid body; in the incremental calculation, each reduction is taken as a certain value of the initial blank height.

## 4. Simulation Results and Analyses

**4.1. Rheological Casting Simulation of Solid-State Processing of Metallic Materials.** A closed die is used to actively load the isothermal extrusion to form a curved metal material. By optimizing the cavity structure of the concave mold, the nondevelopable surface can be directly forged to achieve a net shape or an approximate net shape processing. The simulation results show that the metal flows smoothly, the forming effect is good, and there are no defects such as folding and dissatisfaction with the filling, which proves that the isothermal extrusion forming of curved metal materials is feasible. As shown in Figure 3, when the load is 2 kPa, the vector change of the true radial plastic strain of the working cone is mainly deformed by wall reduction and the amount of wall reduction is evenly distributed; when the load is 3 kPa, the true radial plastic strain has remained unchanged and its value is the largest in the entire forming process because this stage is in the sizing stage of the cold rolling process; when the load is 6 kPa, the residual deformation of the short-side simply supported plate is similar to that of the

free plate, which is greater than the deformation of the short-side fixed plate. It can be seen from the simulated load curve that the deformation process is roughly divided into three stages, namely, the initial deformation, cavity filling, and final forming [13]. In the initial stage of deformation, the load curve rises gently and the deformation force is small; the upward trend of the load curve increases with the increase of displacement, which shows that deformation is becoming more and more difficult.

In the process of multistroke cold rolling of metal materials, the vector change of the true radial plastic strain of the working cone is divided into three stages. From the beginning to the section, the true radial plastic strain gradually increases, and its increasing trend is the largest. This is because at this stage, the working cone is mainly reduced in diameter and the amount of reduction is the largest, while the reduction in wall deformation is not obvious, which is usually called the empty reducing stage, so the value of this stage is relatively small. The radial plastic true strain also gradually increases, but the increasing trend decreases. This is because at this stage the working cone is mainly deformed by wall reduction and the amount of wall reduction is evenly distributed, while the reduction deformation carried out with the wall reduction is relatively large. The true radial plastic strain has remained unchanged and its value is the largest in the entire forming process because this stage is in the sizing stage of the cold rolling process, and the value remains unchanged. The axial length of the outer surface of the rolled tube is greater than the axial length of the neutral layer, and the axial length of the neutral layer is greater than the axial length of the inner surface, which indicates that the degree of anisotropy of the rolled tube ranges from the outer surface to the neutral layer and to the inner surface. The wall-reducing deformation of the material in the final stage is much greater than the diameter-reducing deformation, and the degree of anisotropy of the rolled tube decreases from the outer surface to the neutral layer to the inner surface.

Under different constraint conditions, the magnitude of the residual stress and the residual deformation are different. The boundary constraint of the short-side fixed plate is the strongest and the deformation inside and outside the plane is restricted, the deformation amount is the smallest, but the residual stress is the largest. The residual deformation of the short-side simply supported plate is similar to that of the free plate, which is greater than the deformation of the short-side fixed plate, but the residual stress is smaller. Comparing the calculation results with or without heat preservation after

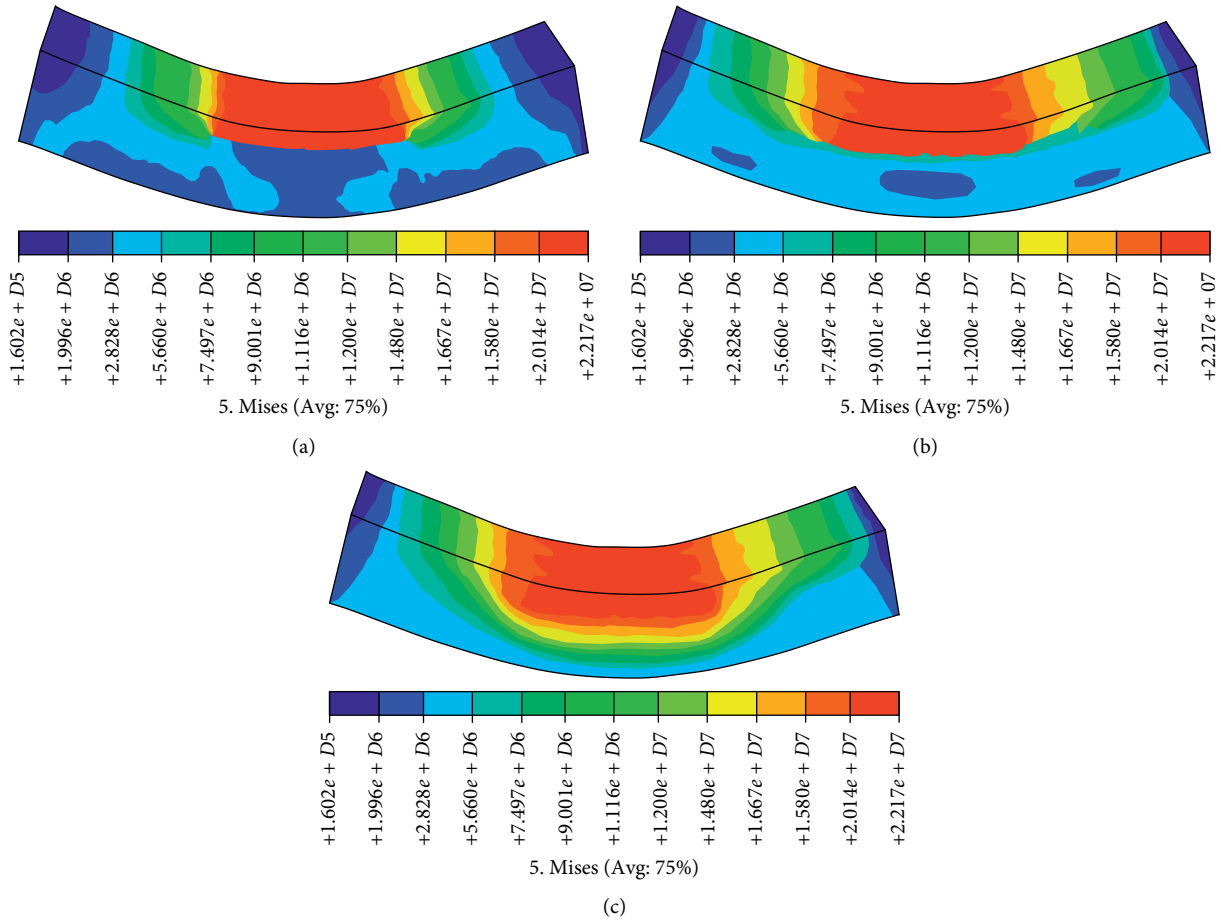


FIGURE 3: 3D finite element numerical simulation results of rheological casting processing of metal material. (a) Load = 2 kPa. (b) Load = 3 kPa. (c) Load = 6 kPa.

processing, whether in terms of residual deformation or residual stress, the performance with heat preservation conditions is significantly better than that without heat preservation. When there is heat preservation, after the material is cooled to the heat preservation temperature, the temperature field is distributed uniformly through heat transfer, and the temperature difference gradient is small. After the heat preservation is removed, the whole material can be uniformly cooled to the on-site ambient temperature; compared with the case of no heat preservation, the material with heat preservation. The degree of uneven cooling is small so that the residual deformation and residual stress are small, indicating that heat preservation measures are indispensable for controlling the residual deformation and residual stress. During the processing and construction of the material shear wall, lateral supports are often set on the two wall panels that need to be processed to limit the effect of the out-of-plane deformation of the material during the processing. After the processing is completed, the support is removed, and the out-of-plane deformation of the material increases [14].

**4.2. Remelting and Heating Simulation of Solid-State Processing of Metal Materials.** In the process of metal deformation, the deformation temperature has a great influence

on the metal deformation. The temperature change during the piercing process has a great influence on the plastic deformation resistance of the metal. When the temperature increases, the plasticity of the metal increases, the deformation resistance weakens, and the metal is easily deformed. On the contrary, the metal deformation resistance increases and the metal are difficult to deform. In addition, temperature is also the key influencing factor for the phase transition of metals, and the phase transition structure is the determining factor for the final mechanical properties of metals and the temperature distribution during the sample rolling process is not uniform. In the direct contact area between the roll and the sample, due to the relative sliding friction between the roll and the sample, it is mainly back slip, causing the temperature in the contact area between the sample and the roll to be higher than the rest (Figure 4). The temperature in the area is tens of degrees higher. The metal equivalent strain in the cross-rolling piercing deformation is the highest near the roll transition cone, and gradually decreases to both sides. The largest deformation occurs in the contact area between the sample and the roll, which corresponds to the stress distribution. During the piercing process, the surface of the plug is subject to high temperature, high pressure, and high shear stress and is prone to wear, melt loss, and steel sticking, which shortens the service

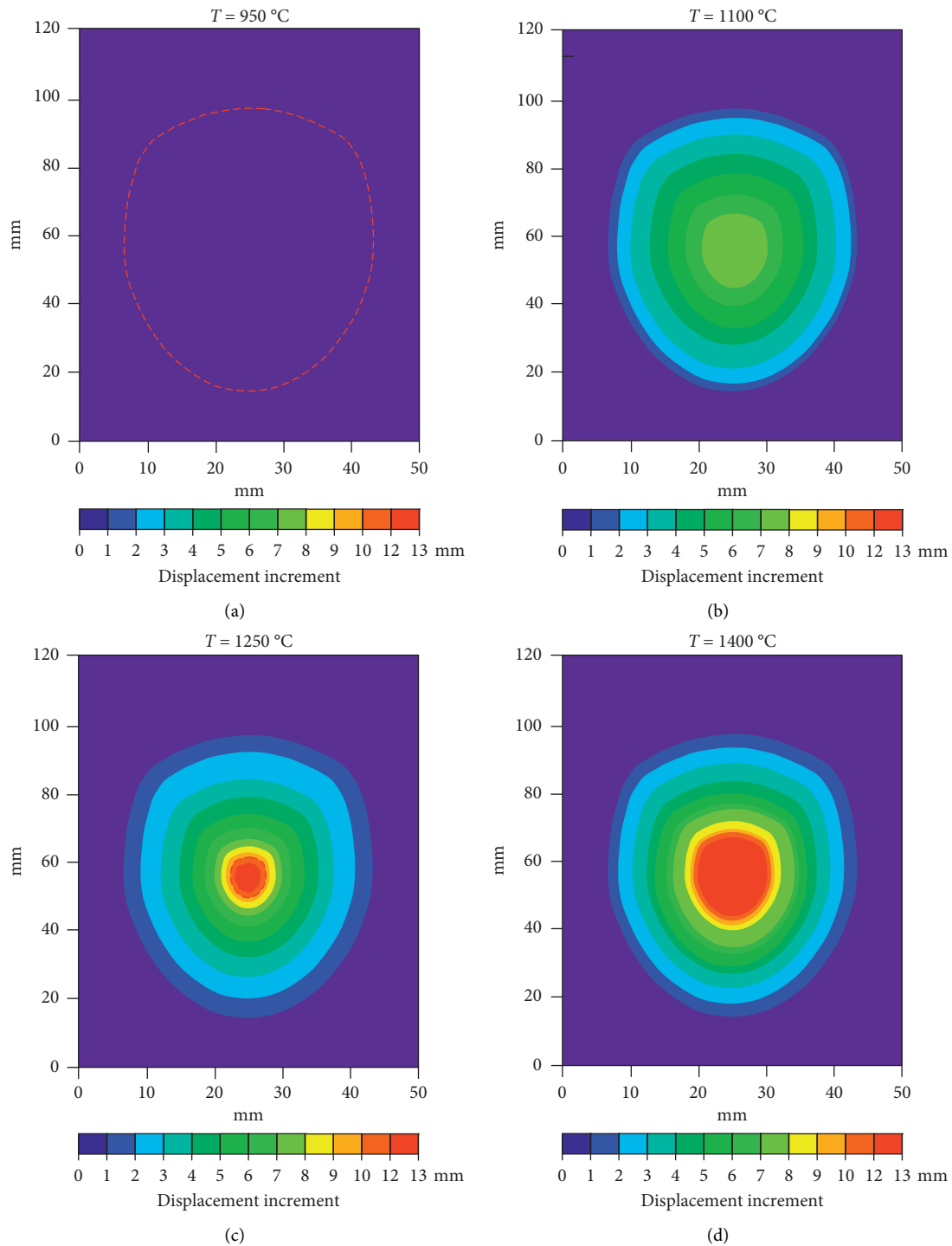


FIGURE 4: Numerical simulation results of displacement increment for mental material at different remelting temperature in solid-state processing.

life of the plug and deteriorates the inner surface quality of the steel pipe. During the cross-rolling and piercing process, the maximum stress on the outer surface of the tube blank occurs in the contact area between the tube blank and the roll and the contact area with the guide plate.

This is because the plastic deformation of the semisolid slurry of the granular structure is mainly achieved by liquid

flow. As the deformation temperature increases, the volume fraction of the liquid phase increases, and the liquid phase is gradually connected into a network to separate the solid particles and the particle rotation and movement are easier to proceed. The solid phase ratio is a key factor affecting the deformation characteristics of the semisolid slurry. Generally speaking, the longer the opening and drawing time, the



closer the position of the crack is to the lower opening of the mold, that is, the length of the retained layer in the mold becomes longer with the extension of the opening and drawing time [15]. The greater the heat flux density, the closer the crack position is to the lower mouth of the crystallizer, that is, the length of the retention layer in the crystallizer increases as the heat flux density increases. The greater the heat flux density, the closer the crack position is to the lower mouth of the crystallizer, that is, the length of the retention layer in the crystallizer increases as the heat flux density increases (Figure 5). The longer the opening time or the greater the heat flux density, the greater the thickness of the solidified shell at the crack, which is consistent with common sense. From the perspective of the corresponding relationship between the crack position and the crack thickness, in general, the lower the crack position, the thicker the thickness of the solidified shell at the crack. If a crack occurs due to excessive drawing resistance, the position of the crack will be lower. Figure 5 concludes that during repeated grinding and deformation, since it is a new deformation based on the plastic deformation of the last pass, the surface metal has undergone work hardening after plastic deformation, so the deformation resistance is greater.

The finite element analysis of solid-state rolling is a highly nonlinear problem. Considering material nonlinearity, boundary nonlinearity, and temperature nonlinearity, the use of simple geometric models is a means to make research progress. In order to study the change of solid state during rolling, two rolling methods, flat rolling and closed pass rolling, were used; in flat rolling, the solid slurry is simplified into a round bar. In the pass rolling, under the restriction of the pass and side sealing tools, the solid slurry has better fluidity and will form a small hole above the hole; therefore, the solid slurry can be simplified into cuboids. Because flat rolling has no restriction on the side of the rolled piece, the horizontal extension is large. At the same time, due to the friction force of the rolling, the contact part of the rolling piece head and the rolling piece extends in the longitudinal direction to form a spiral depression [16]. The downward pulling force causes the remaining part of the rolled piece to shrink and become thinner. Because the solid material has low flow stress and good fluidity, it can be rolled with a large amount of deformation, but at the same time, it is easy to cause uneven stress distribution on the cross-section. The stress distribution on the cross-section of the deformation zone is uneven, decreasing from the core to the side on the same cross-section; the longitudinal changes of the stress gradient on the surface of the deformation zone are inconsistent.

## 5. Discussion

*5.1. Thixoforming Simulation of Solid-State Processing of Metallic Materials.* From the relationship between volume change rate and pressing time under different load conditions, it can be seen that, in the initial stage of pressing, the volume changes more drastically. This is mainly due to the rapid decrease of the void density between the compact particles and the gradual compaction of the metal material

particles. As the pressing progresses, the volume change trend is relatively gentle. At this stage, the metal material particles are rearranged and plastically deformed. From the comparison between the experimental value and the simulated value, it can be seen that the change of the average density goes through three stages with the increase of the load: the rapid growth stage, the gentle growth stage, and the saturation stage. The density distribution is in a stepped layered distribution that gradually increases from bottom to top in parallel to the movement direction of the die punch (Figure 6). The maximum and minimum relative density of the compact are located at the end corners of the upper and lower surfaces, respectively, inside the press. The density distribution is more uniform than the outer surface, and the density distribution is a contour surface, which indicates that the radial flow and axial flow of the metal material are in a balanced state on this surface. The density distribution is layered parallel to the movement direction of die punching, and the density distribution on the outer surface is basically gradually increasing from bottom to top, and is basically parallel to the top surface. In the interior of the press, the density distribution is stepped. Gradually, except for the upper and lower corners, the density distribution inside the pressing piece is more even than the outer surface [17].

The flow velocity of the melt in the longitude direction is much larger than the axial casting velocity. The electromagnetic field in the velocity field presents a distribution and reaches its maximum value on the middle surface of the stator. The equivalent flow trajectory length of all solidified particles increases with the progress of casting, and the residence time of solid particles in the solidification zone is also significantly different. This method can accurately process free surfaces, effectively track the deformation history of fluid masses, conveniently realize grid optimization, obtain local accurate physical images, and can handle larger grid deformations. Numerical simulation technology can not only simulate the changes in the flow field, temperature field, stress field, and microstructure of the metal processing process but also predict the formation of defects in the metal processing process. Therefore, it has control over the understanding and analysis of the solid metal forming process, and optimization of process parameters can play an effective guiding role for rheological casting, secondary heating, and near-net thixoforming of parts. However, solid metal forming is a solid-liquid two-phase flow process and has importance, and its deformation resistance or apparent viscosity is not only related to the shear rate but also closely related to time. Therefore, the numerical simulation method must be able to track the fluid deformation process.

Under the same reduced pressure condition, the surface of the plate is milled by single milling and repeated milling. The comparison of the effective strain shows that the cumulative effect of repeated milling on the effective strain is very obvious [17]. The reason for this result is that the deformation of the metal material during processing consists of two parts: elastic deformation and plastic deformation. The elastic deformation is restored after each grinding process, and the plastic deformation remains in the

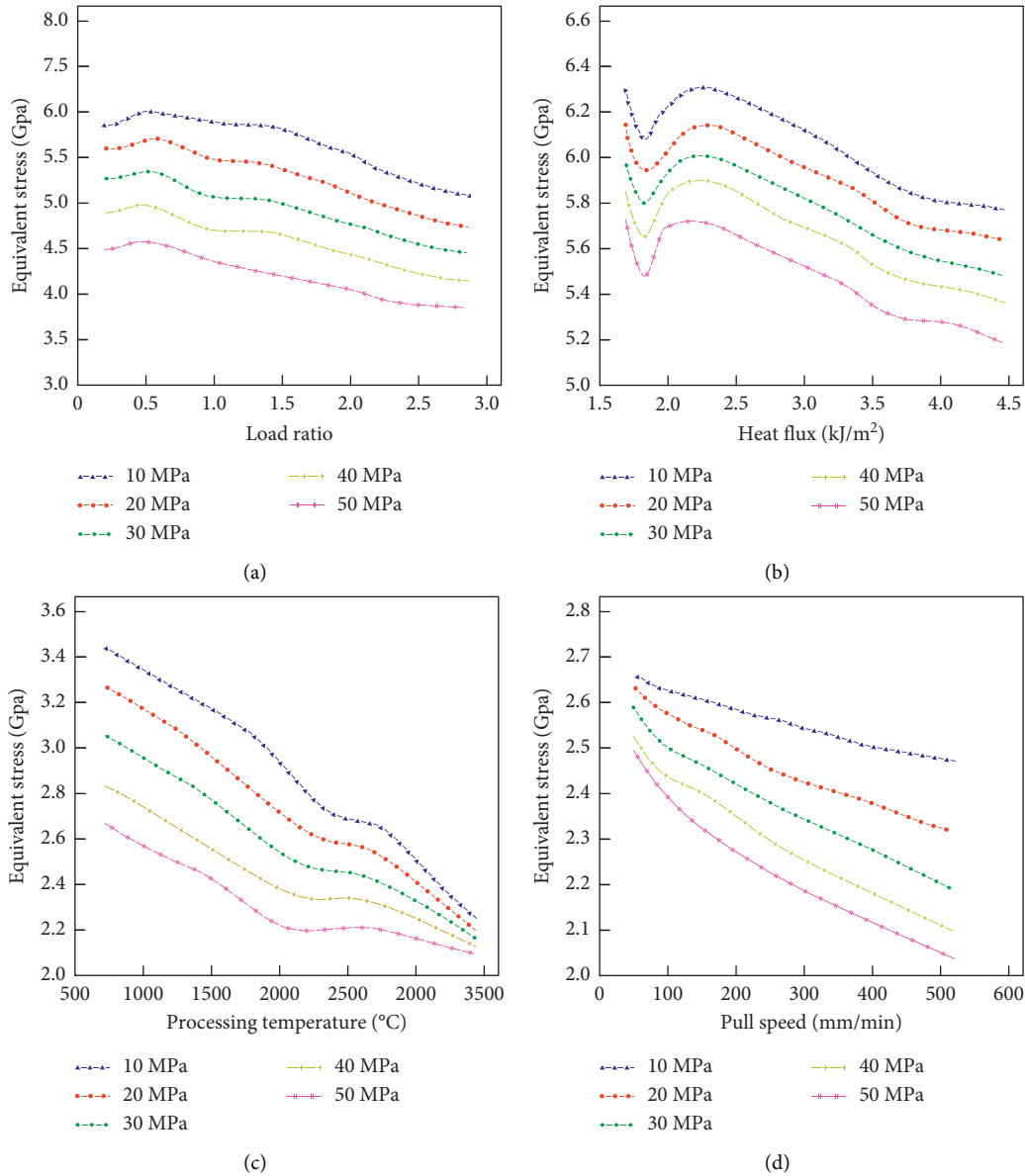


FIGURE 5: Relationship between equivalent stress and load ratio. (a) Heat flux. (b) Processing temperature. (c) Pull speed. (d) Different processing effects.

deformation area. During repeated grinding and deformation, since it is a new deformation based on the plastic deformation of the last pass, the surface metal has undergone work hardening after plastic deformation, so the deformation resistance is greater. In order to continue the deformation, the plate must be the material exerts greater compressive stress, which is also the reason why more effective strain can be obtained by multipass grinding under the same lower pressure condition. With the increase of milling passes, the maximum effective strain gradually increases. When the number of milling passes exceeds the second, even if the number of milling passes increases, the maximum effective strain does not increase much, indicating that the strengthening effect produced by strong plastic deformation tends to be saturated. No matter which milling method is used, the effective strain increases with the

increase of milling times, and the curve is close to the linear change law.

*5.2. Rotary Piercing Simulation of Solid-State Processing of Metal Materials.* The alloy ingot with the structure is heated and sheared during the transportation process through a heating source and a special spiral propulsion system so that it has importance and is injected into the mold cavity to form injection molding which is essentially a kind of rheological technology. As shown in Figure 7, in stage (1), wall thickness changes everywhere are not gradually and slowly, and the change at the beginning quickly reaches the final change result; in stage (2), the bending part of the tube gradually moves to the rear end of the tube as the bending angle increases; in stage (3), the wall thickness is still the initial

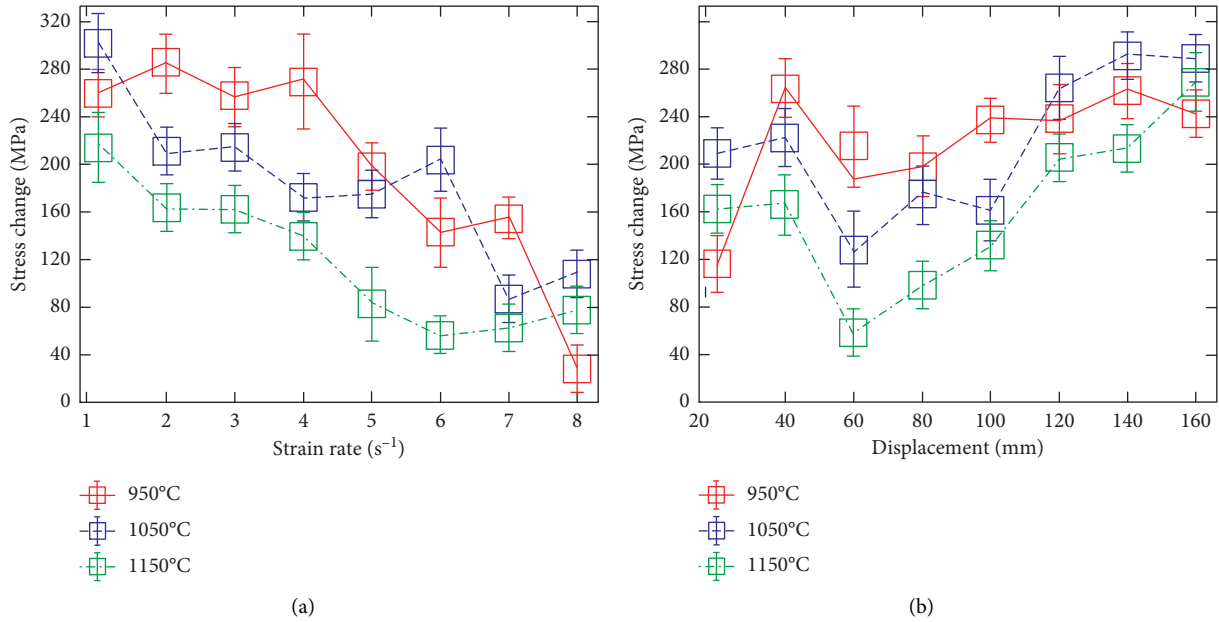


FIGURE 6: Relationship between stress change and strain rate. (a) Displacement. (b) Mental material in the solid-state processing.

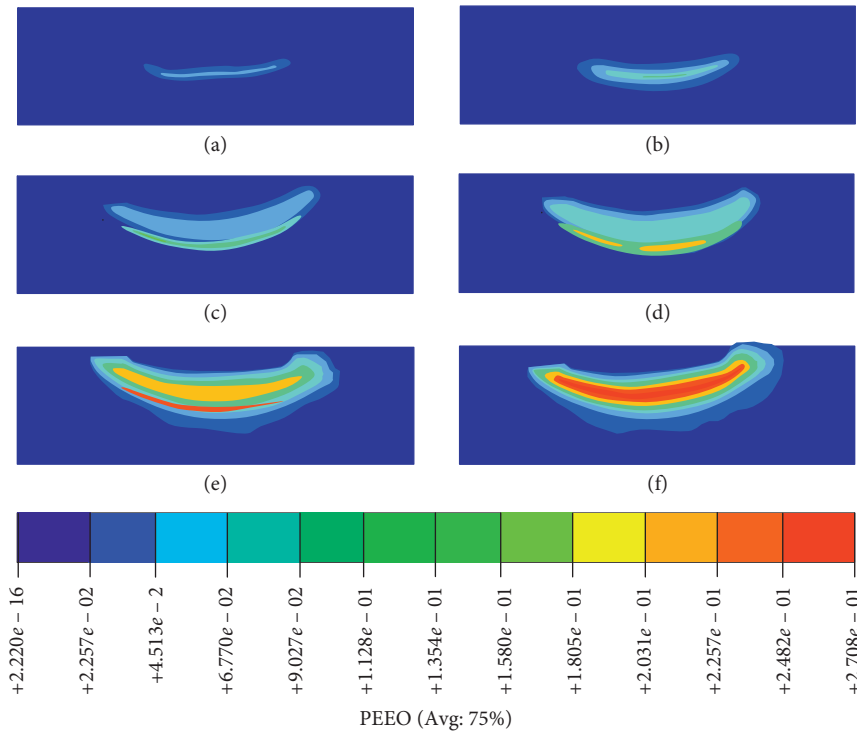


FIGURE 7: Rotary piercing simulation results of solid-state processing of metal materials. (a) Stage (1). (b) Stage (2). (c) Stage (3). (d) Stage (4). (e) Stage (5). (f) Stage (6).

value; as the bending angle increases, the deformation zone moves to this point; in stage (4), tensile and compressive stresses are generated on the pipe wall, resulting in a change in the wall thickness; in stage (5), no bending deformation occurs, and the wall thickness at this time remains basically unchanged; in stage (6), the inner concave edge of the elbow is compressed to increase the wall thickness, while the outer

convex edge is stretched to reduce the wall thickness. The total change of the equivalent stress value during the bending deformation process of the pipe fitting increases with the increase of the bending angle, and the stress value of the inner concave side pipe wall of the curved pipe fitting is larger than that of the outer convex side pipe wall. The inner wall is subjected to compressive stress and the outer wall is

subjected to compressive stress. The equivalent strain value of the inner pipe wall is larger than the outer pipe wall. In the bending process, due to the small bending radius and large bending deformation, the inner concave edge of the elbow is compressed to thicken the inner wall, while the outer convex edge is stretched to make the outer wall thinner [18].

When the strain rate is low, driven by the pore pressure gradient, the liquid seeps out of the free surface, and the resulting pore pressure is much smaller than the total macroscopic pressure. At the other extreme, if the deformation is fast, the liquid does not have enough time to migrate to low pressure zone and causes the accumulation of pressure and the deformation in this case is nondraining [19]. In the nondraining deformation, the increased pressure reduces the pressure component of the equivalent stress, resulting in a reduction in the three-way equivalent stress because the reduction of the cohesive force is directly related to the partial stress of the applied stress. In the case of no drainage, the solid skeleton is destroyed, which easily causes deformation concentration. Without ignoring the pressure of the liquid, there is a lack of understanding of the behavior of metals in liquid-filled porous media both theoretically and experimentally. Because these relationships are not only theoretical, technological processes such as continuous casting and solid forging are likely to generate a large amount of liquid pressure. The accurate simulation of these processes requires the constitutive equation of the liquid-solid mixture, and the contribution of the liquid to the deformation mechanism of the mixture needs to be accurately described. The asymmetry of the metal solid down-compression deformation behavior and the use of liquid compression experiment, nondraining liquid compression experiment, and shear experiment verified the correctness of the model [20].

In the actual forming process, the size range of different partitions is constantly changing, and the related areas are also constantly transformed. When they are in the stable deformation stage, the metal volume from the force transmission zone into the deformation zone and transfer from the deformation zone to the deformed volume of the zone is equal. With the progress of the deformation process, the force-transmitting zone keeps decreasing; the deformed zone keeps increasing, while the size of the deformed zone and the numerical distribution of stress remain unchanged [21]. During the deformation process, the degree of deformation, the diameter, thickness, length of the tube blank, and the adjustment of the structural parameters of the rotary wheel will all have a direct impact on the deformation state. When the feed depth is constant, the deformation resistance is reduced when the working angle is small, which is more conducive to metal deformation, but on the contrary, the distance of metal flow increases, and the deformation zone is correspondingly elongated, which increases the contact area between the blank and the surface of the rotary wheel. When the working angle of the rotary wheel is large, the contact area between the blank and the rotary wheel is reduced, and the deformation of the metal is more difficult, resulting in an increase in the deformation force. In the process of necking forming, the deformation zone is mainly affected by radial

and axial compressive stresses, which reduces the diameter and increases the wall thickness and length. For thin-walled cylinders, when the radial compressive stress is too large, it is easy to cause wrinkling and instability in the deformed thin-walled area.

## 6. Conclusions

This paper conducted finite element model's geometric assumptions, material selection, element division, model establishment, parameter selection, and initial and boundary condition determination and simulated and analyzed rheological casting, remelting heating, thixoforming, and rotary piercing processes of metal materials. From the relationship between volume change rate and pressing time under different load conditions, it can be seen that, in the initial stage of pressing, the volume changes more drastically. This is mainly due to the rapid decrease of the void density between the compact particles and the gradual compaction of the metal material particles. At this stage, the metal material particles are rearranged and plastically deformed and the density distribution is in a stepped layered distribution that gradually increases from bottom to top in parallel to the movement direction of the die punch. The maximum and minimum relative density of the compact are located at the end corners of the upper and lower surfaces, respectively, inside the press. The density distribution is more uniform than the outer surface, and the density distribution is a contour surface, which indicates that the radial flow and axial flow of the metal material are in a balanced state on this surface. The results show that the 3D finite element numerical method can not only simulate various processes of flow field, temperature field, stress field, and microstructure in solid-state processing but also can provide a reliable basis for effectively obtaining a reasonable description and finding a more optimized design plan for metal material processing in a short time, which plays an important role in understanding and analyzing solid metal forming process, controlling and optimizing process parameters, guiding and mastering rheological casting, and secondary heating and rotary piercing of metal materials.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

The authors acknowledge the Project of Henan Science and Technology (192102210164) and Doctoral Research Start-up Fund of Henan Institute of Technology (KQ1848).

## References

- [1] J. Martín, D. Zhao, T. Lenz, I. Katsouras, D. M. de Leeuw, and N. Stingelin, "Solid-state-processing of  $\delta$ -PVDF," *Materials Horizons*, vol. 4, no. 3, pp. 408–414, 2017.
- [2] A. R. Yoo, S. A. Yoon, Y. S. Kim, J. Sakamoto, and H. C. Lee, "A comparative study on the synthesis of Al-doped  $\text{Li}_6\text{La}_3\text{Zr}_2\text{O}_{12}$  powder as a solid electrolyte using sol-gel synthesis and solid-state processing," *Journal of Nanoscience and Nanotechnology*, vol. 16, no. 11, pp. 11662–11668, 2016.
- [3] J. Suh, S. Shin, and D. Bae, "Electrical properties of polytetrafluoroethylene/few-layer graphene composites fabricated by solid-state processing," *Journal of Composite Materials*, vol. 51, no. 18, pp. 2565–2573, 2017.
- [4] V. Patel, W. Li, A. Vairis, and V. Badheka, "Recent development in friction stir processing as a solid-state grain refinement technique: microstructural evolution and property enhancement," *Critical Reviews in Solid State and Materials Sciences*, vol. 44, no. 5, pp. 378–426, 2019.
- [5] M. Nagaral, S. Kalgudi, V. Auradi, and S. A. Kori, "Mechanical characterization of ceramic nano B4C- Al2618 alloy composites synthesized by semi solid state processing," *Transactions of the Indian Ceramic Society*, vol. 77, no. 3, pp. 146–149, 2018.
- [6] R. Kroon, A. I. Hofmann, L. Yu, A. Lund, and C. Müller, "Thermally activated in situ doping enables solid-state processing of conducting polymers," *Chemistry of Materials*, vol. 31, no. 8, pp. 2770–2777, 2019.
- [7] D. Desai, H. Sandhu, N. Shah et al., "Selection of solid-state plasticizers as processing aids for hot-melt extrusion," *Journal of Pharmaceutical Sciences*, vol. 107, no. 1, pp. 372–379, 2018.
- [8] H. Takizawa, "Survey of new materials by solid state synthesis under external fields: high-pressure synthesis and microwave processing of inorganic materials," *Journal of the Ceramic Society of Japan*, vol. 126, no. 6, pp. 424–433, 2018.
- [9] E. V. Dedyeva, D. V. Zaitsev, E. A. Lukina et al., "Effect of barothermal processing on the solid-state formation of the structure and properties of 16 at % Si-Al hypereutectic alloy," *Inorganic Materials*, vol. 54, no. 2, pp. 125–132, 2018.
- [10] R. Duliński, A. Starzyńska-Janiszewska, L. Byczyński, and U. Błaszczak, "Myo-inositol phosphates profile of buckwheat and quinoa seeds: effects of hydrothermal processing and solid-state fermentation with *Rhizopus oligosporus*," *International Journal of Food Properties*, vol. 20, no. 9, pp. 2088–2095, 2017.
- [11] A. C. Oliveira, G. M. Amorim, J. A. G. Azevêdo, M. G. Godoy, and D. M. G. Freire, "Solid-state fermentation of co-products from palm oil processing: production of lipase and xylanase and effects on chemical composition," *Biocatalysis and Bio-transformation*, vol. 36, no. 5, pp. 381–388, 2018.
- [12] D. M. Vasa and P. L. D. Wildfong, "Solid-state transformations of ribavirin as a result of high-shear mechanical processing," *International Journal of Pharmaceutics*, vol. 524, no. 1–2, pp. 339–350, 2017.
- [13] Y. Yamamoto, Y. Iriyama, and S. Muto, "Union operation image processing of data cubes separately processed by different objective filters and its application to void analysis in an all-solid-state lithium-ion battery," *Microscopy*, vol. 65, no. 2, pp. 191–198, 2016.
- [14] R. Ashiri, "On the solid-state formation of  $\text{BaTiO}_3$  nanocrystals from mechanically activated  $\text{BaCO}_3$  and  $\text{TiO}_2$  powders: innovative mechanochemical processing, the mechanism involved, and phase and nanostructure evolutions," *RSC Advances*, vol. 6, no. 21, pp. 17138–17150, 2016.
- [15] A. Mazumdar and H. Maumdar, "Bio-processing of banana peel for alpha amylase production by *Aspergillus oryzae* employing solid state fermentation," *The Clarion- International Multidisciplinary Journal*, vol. 7, no. 1, pp. 36–42, 2018.
- [16] M. Dagnaw and B. Andualem, "Solid state fermentation of keratinolytic proteases production using *Bacillus* spp. isolated from hair and mud sample of traditional leather processing ponds in North Gondar, Ethiopia," *Journal of Medicinal Plants Studies*, vol. 7, no. 4, pp. 127–138, 2019.
- [17] I. Aharonovich, D. Englund, and M. Toth, "Solid-state single-photon emitters," *Nature Photonics*, vol. 10, no. 10, pp. 631–641, 2016.
- [18] W. T. Hong, M. Risch, K. A. Stoerzinger, A. Grimaud, J. Suntivich, and Y. Shao-Horn, "Toward the rational design of non-precious transition metal oxides for oxygen electrocatalysis," *Energy & Environmental Science*, vol. 8, no. 5, pp. 1404–1427, 2015.
- [19] Y. Wang, W. D. Richards, S. P. Ong et al., "Design principles for solid-state lithium superionic conductors," *Nature Materials*, vol. 14, no. 10, pp. 1026–1031, 2015.
- [20] D. P. Dubal, N. R. Chodankar, D.-H. Kim, and P. Gomez-Romero, "Towards flexible solid-state supercapacitors for smart and wearable electronics," *Chemical Society Reviews*, vol. 47, no. 6, pp. 2065–2129, 2018.
- [21] C. Sun, J. Liu, Y. Gong, D. P. Wilkinson, and J. Zhang, "Recent advances in all-solid-state rechargeable lithium batteries," *Nano Energy*, vol. 33, no. 33, pp. 363–386, 2017.

## Research Article

# Accurate Recognition and Simulation of 3D Visual Image of Aerobics Movement

Wenhua Fan<sup>1,2</sup> and Hyun Joo Min<sup>2</sup> 

<sup>1</sup>Sports Institute, Jiaying University, Meizhou 514015, China

<sup>2</sup>Sports Institute, Korea Gangneung-Wonju National University, Gangneung 25457, Republic of Korea

Correspondence should be addressed to Hyun Joo Min; 201101029@jyu.edu.cn

Received 7 September 2020; Revised 13 October 2020; Accepted 27 October 2020; Published 12 November 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Wenhua Fan and Hyun Joo Min. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The structure of the deep artificial neural network is similar to the structure of the biological neural network, which can be well applied to the 3D visual image recognition of aerobics movements. A lot of results have been achieved by applying deep neural networks to the 3D visual image recognition of aerobics movements, but there are still many problems to be overcome. After analyzing the expression characteristics of the convolutional neural network model for the three-dimensional visual image characteristics of aerobics, this paper builds a convolutional neural network model. The model is improved on the basis of the traditional model and unifies the process of aerobics 3D visual image segmentation, target feature extraction, and target recognition. The convolutional neural network and the deep neural network based on autoencoder are designed and applied to aerobics action 3D visual image test set for recognition and comparison. We improve the accuracy of network recognition by adjusting the configuration parameters in the network model. The experimental results show that compared with other simple models, the model based on the improved AdaBoost algorithm can improve the final result significantly when the accuracy of each model is average. Therefore, the method can improve the recognition accuracy when multiple neural network models with general accuracy are obtained, thereby avoiding the complicated parameter adjustment process to obtain a single optimal network model.

## 1. Introduction

Aerobics is a kind of gymnastics, it is a kind of gymnastics that can change people's physical and psychological feelings. As an independent event, the sport joined the International Gymnastics Federation in September 1994. In recent years, with the call of the global fitness craze and happy sports, the recognition of aerobics has become more and more widespread. The development of fitness aerobics and competitive aerobics conforms to the trend of the times [1, 2]. The essence of aerobics is to let people enjoy the feeling of joy through the beauty of athletes' wonderful performances. This is a manifestation of high-grade spiritual civilization and wealth. It is precisely because of the fact that it attracts more people to participate.

The three-dimensional visual image of aerobics is the basis of human perception of the world and provides people with rich and diverse information [3]. People have a strong ability

to recognize the three-dimensional visual images of aerobics movements and are not even restricted by the sensory channel [4, 5]. People can not only recognize words with their eyes, but also when they write on the back of a person. Through the study of eye activity during the 3D visual image recognition of aerobics, it is found that changing the distance of the 3D visual image of aerobics and the position on the sensory organs will cause the size and shape of the 3D visual image of aerobics to change on the retina [6]. In the current 3D visual image recognition system for aerobics movements, the recognition of complex 3D visual images of aerobics movements can be realized mainly through different levels of information processing [7, 8]. For the familiar three-dimensional visual image of aerobics movements, once you have mastered its main features, you can use this main feature as a recognition unit without paying attention to its details [9]. The 3D visual image recognition of aerobics movements is an important application of artificial intelligence. Its main function is to compile

a computer program that imitates the human brain to recognize the three-dimensional visual image of aerobics movements, process the three-dimensional visual image of aerobics movements, and extract effective information from it to realize the recognition of people or things [10, 11]. The development process of 3D visual image recognition of aerobics movements has mainly experienced three stages: text recognition, 3D visual image information processing, and recognition of aerobics movements and object recognition [12]. Now some of the relatively mature aerobics 3D visual image recognition technologies have been applied to the commercial field [13]. Due to the diversity of problems in the process of 3D visual image recognition of aerobics movements, the methods of 3D visual image recognition of aerobics movements can only adopt different methods for specific problems [14]. Therefore, many recognition systems require a lot of research to improve the performance of breakthroughs on specific issues (such as improving the efficiency of recognition and reducing the time required for system training). The method of machine learning is a method that can get better recognition results on different recognition problems [15]. Relevant scholars have adopted sparse coding that is more suitable for the three-dimensional visual image representation of aerobics to identify and recognize the target object, instead of vector quantization coding that has a large loss of accuracy [16]. The 3D visual image recognition technology of aerobics has been continuously improved. But we have to say that there are still many unresolved problems. The quality of the 3D visual image of aerobics action determines the effect of 3D visual image recognition of aerobics action, and the effect of 3D visual image recognition of aerobics action of poor quality is not good [17, 18]. There are still many issues that need to be studied further [19, 20]. In general, the 3D visual image recognition of aerobics movements is an extremely challenging task. It is difficult to achieve the desired effect with only one method. How to improve the accuracy of the 3D visual image recognition of aerobics movements and reduce the algorithm complexity of the system and how to make it practical are all worth studying.

This paper improves on the mainstream convolutional neural network model VGG and builds a new CNN network model, which is called N-VGG. Combined with this model, a simple aerobics action three-dimensional visual image target positioning and recognition model is constructed. The activation function used in the traditional convolutional neural network model is analyzed, and a new activation function is introduced into the N-VGG network model. In the N-VGG network model, the spatial pyramid pooling technology (SPP) is introduced to improve the recognition accuracy of the model. This paper conducts multiple comparative experiments on convolutional neural networks and deep neural networks and shows that a single neural network can be adjusted with different parameters to improve its accuracy, but its parameter adjustment process is complicated and random. By comparing the experimental results of convolutional neural network and deep neural network fusion in single and different methods, it can be concluded that the model fusion method based on the improved AdaBoost algorithm can improve the neural

network's ability to improve the 3D visual image of aerobics movements to a certain extent.

The rest of this article is organized as follows. Section 2 studies the 3D visual image recognition technology and related algorithms of aerobics. Section 3 designs a convolutional neural network recognition model. In Section 4, simulation experiments and result analysis are carried out. Section 5 summarizes the full text.

## 2. Aerobics 3D Visual Image Recognition Technology and Related Algorithms Research

*2.1. 3D Visual Image Processing of Aerobics.* Aerobics 3D visual image processing refers to the use of a computer to process the 3D visual image of aerobics to be recognized to meet the subsequent needs of the recognition process. It is mainly divided into two types: aerobics 3D visual image preprocessing and aerobics 3D visual image segmentation. The preprocessing of 3D visual image of aerobics mainly includes restoration of 3D visual image of aerobics and transformation of 3D visual image of aerobics. Its main purpose is to remove the interference and noise in the 3D visual image of aerobics and enhance the 3D vision of aerobics. The useful information in the image improves the detectability of the target object. At the same time, due to the real-time requirements of the three-dimensional visual image processing of aerobics, it is necessary to re-encode and compress the three-dimensional visual image of aerobics to reduce the complexity and complexity of subsequent algorithms. The segmentation of aerobics 3D visual image is the process of segmenting the 3D visual image of aerobics to be recognized into several subregions. The features of each region have obvious differences, and the internal features of each region have certain similarities.

The method based on edge segmentation is to segment the 3D visual image of aerobics by detecting the area where the gray value of the pixel changes suddenly, or the place where the texture structure changes suddenly. The edge is usually located at the connection between two different areas. Because in a 3D visual image of aerobics action, the gray values of different areas are often different, and there will be obvious gray discontinuities at the junction of the two areas, that is, the edge. Because the gray value of each pixel at the edge is discontinuously distributed, differential or second-order differential can be used for detection. They calculate the first-order differential of the gray value of each pixel distributed in the edge area; then the pixel corresponding to the place where its extreme value appears is the edge point of the three-dimensional visual image of aerobics action. The gray value of a point is calculated for the second-order differential, and the pixel point with a differential value of zero is also the edge point of the three-dimensional visual image of aerobics. Therefore, the edge detection of the three-dimensional visual image of aerobics can be performed by the method of differential operators.

Because the Roberts differential detection operator uses an even number of templates, the gradient amplitude value at  $(x, y)$  is actually the value at the intersection shown in Figure 1, which is offset by the pixel from the real position, so

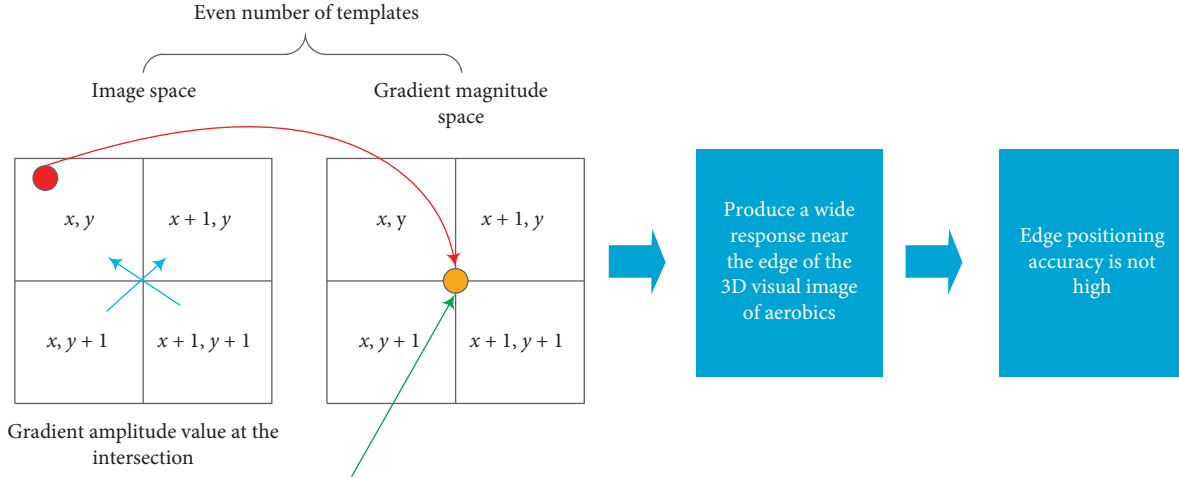


FIGURE 1: Schematic diagram of Roberts operator algorithm.

this method will produce a wider response near the edge of the 3D visual image of aerobics, and the edge positioning accuracy is not high.

### 2.2. Feature Extraction of the 3D Visual Image of Aerobics.

The feature extraction of the 3D visual image of aerobics is the key step in the three steps of 3D visual image recognition of aerobics. It refers to the 3D vision of aerobics that remains unchanged after the interference of factors such as scaling, translation, scale transformation, or illumination. Image information is extracted as features, and the three-dimensional visual image of aerobics is abstracted into some concrete mathematical representations or vector descriptions. The feature extraction of aerobics action three-dimensional visual image can be divided into global features and local features according to the different extraction ranges. The global features of the three-dimensional visual image of aerobics refers to all the features that describe the overall information of the three-dimensional visual image of aerobics, such as the color feature, shape feature, and texture feature of the three-dimensional visual image of aerobics. The local features refer to the partial features of the 3D visual image of aerobics movements in a specific area. It is only a representation of the local features in a certain area in the 3D visual image of aerobics movements. It is only suitable for matching the 3D visual image of aerobics movements.

The color histogram is the most widely used color feature in the 3D visual image retrieval of aerobics. It mainly uses the statistical distribution and basic tones of different colors in the entire 3D visual image of aerobics. It does not care about each color in aerobics. Each aerobics action 3D visual image can find its unique color histogram, so it is especially suitable for describing the 3D visual image of aerobics action that cannot be automatically segmented. The color histogram can be divided into two description methods: feature statistical histogram and feature cumulative histogram.

Suppose the total number of pixels in the three-dimensional visual image  $P$  of aerobics action with feature value  $x_i$  is  $s(x_i)$ , and  $N$  is the total number of pixels in the

three-dimensional visual image of aerobics action. The proportion of the total pixels in the 3D visual image is

$$h(x_i) = \frac{s(x_i)}{\sum_j s(x_j)}. \quad (1)$$

That is to say, the feature histogram with feature  $x$  in the three-dimensional visual image  $P$  of aerobics can be expressed as

$$H(P) = [h(x_1), h(x_2), h(x_3), \dots, h(x_{n-1}), h(x_n)]. \quad (2)$$

Among them,  $n$  represents the number of features  $x$ . The cumulative histogram with feature  $x$  in the three-dimensional visual image  $P$  of aerobics movements can be expressed as

$$u(P) = [u(x_1), u(x_2), u(x_3), \dots, u(x_{n-1}), u(x_n)]. \quad (3)$$

Therefore, the feature statistical histogram and feature accumulation histogram of the three-dimensional visual image of aerobics can be regarded as two one-dimensional discrete functions, that is, to obtain the probability distribution of a certain feature in the three-dimensional visual image of aerobics. The feature histogram is the probability distribution of the grayscale of the three-dimensional visual image of aerobics.

**2.3. Discriminant Deep Recognition Structure.** The main function of discriminative depth structure is to distinguish pattern recognition and describe the posterior probability distribution of data. Because CNN contains multiple levels of network structure, it can ensure the minimization of preprocessed data, rely on shared weights to reduce the complexity of the model structure, and use spatial relationships to reduce the number of parameters. It is a learning algorithm that can perform better training. Unlike the DBN network, it is a discriminative deep structure. The CNN algorithm model structure is mainly composed of convolutional layer, pooling layer, and fully connected layer, as shown in Figure 2. Using the CNN model structure to



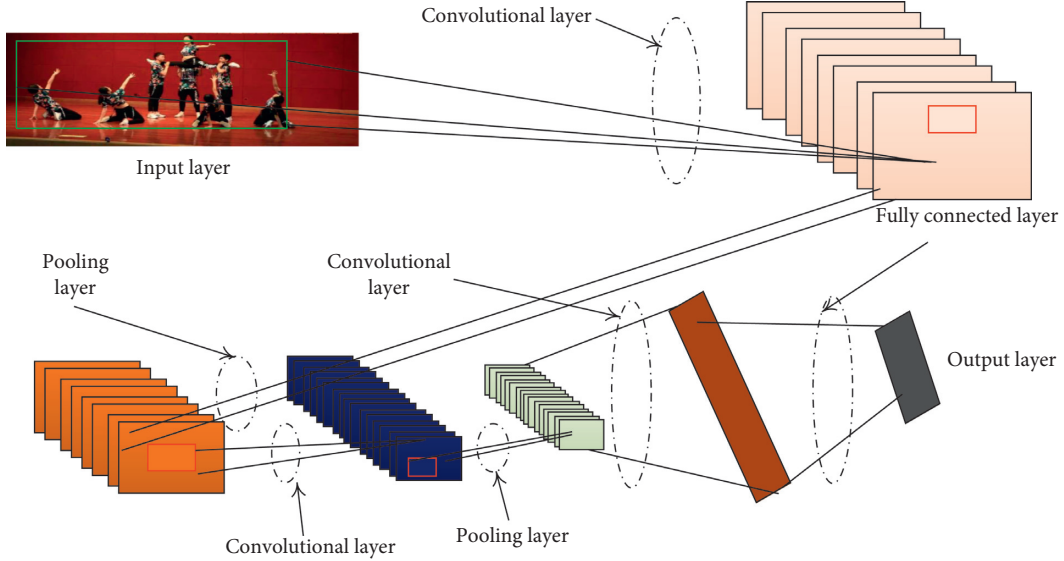


FIGURE 2: Model diagram of 3D visual image recognition of aerobics based on CNN.

process the three-dimensional visual image of aerobics can make full use of the two-dimensional structure of the three-dimensional visual image data of aerobics to perform good feature extraction on the three-dimensional visual image of aerobics. Therefore, compared with other deep learning models, CNN has a higher accuracy in the recognition of 3D visual images of aerobics.

CNN first convolves the input aerobics action three-dimensional visual image with the convolution kernel of the convolution layer and the bias that can be added. The basic mathematical expression of convolution in calculus is

$$S(t) = \int w(a) \cdot x(t-a) da. \quad (4)$$

The discrete form is expressed as

$$s(t) = \sum_a w(a) \cdot x(t-a). \quad (5)$$

For the two-dimensional convolution operation in the CNN network, the mathematical expression is

$$s(i, j) = \sum_m \sum_n w(m, n) \cdot x(i+m, j+n). \quad (6)$$

Among them,  $W$  is the convolution kernel in the convolution layer, and  $X$  is the input data. If  $X$  is a two-dimensional input matrix,  $W$  is also a two-dimensional matrix. If  $X$  is a multidimensional vector, then  $W$  is also a multidimensional vector. When the convolution layer performs a convolution operation on the input aerobics action three-dimensional visual image, it is the convolution kernel matrix and the input aerobics action three-dimensional visual image. The elements in each position of the different local matrix are multiplied and then added, but the operation is a kind of linear operation, while the neural network requires that each neuron in the network can adapt to complex nonlinear operations, which requires the introduction of nonlinear factors, namely, activation functions.

In the CNN model, there are multiple convolution kernels in the convolutional layer. After each convolution kernel performs the convolution operation, the feature map of the input aerobics action 3D visual image can be obtained. These feature maps share the same weight matrix and bias. After the convolution results are averaged, the 3D visual image of the characteristic aerobics action is output, namely,

$$Y_j = \frac{\sum_{i=1}^N X_i \cdot W_j}{N}, \quad (7)$$

where  $X_i$  represents the feature map extracted by the  $i$ th input,  $W_j$  represents the weight matrix of the  $j$ th convolution kernel,  $N$  represents the total number of input feature maps, and  $Y_j$  represents the output  $j$ th feature map.

After the 3D visual image of aerobics moves through the convolutional layer to obtain the feature map of the 3D visual image of the aerobics movement, it is sent to the pooling layer to aggregate and count each feature. The pooling layer is mainly responsible for the aggregation of the three-dimensional visual images of the characteristic aerobics actions, using statistical features to obtain the average or maximum value of the three-dimensional visual images of the aerobics actions in the characteristic area, and removing irrelevant feature samples to further reduce the number of parameters. The use of pooling operations on the continuous areas in the 3D visual image of aerobics can reduce the effect of translation, rotation, and other factors on the 3D visual image of aerobics and prevent overfitting.

The fully connected layer is the “recognizer” of the entire CNN network. Its main function is to map the feature representation after operations such as convolution and pooling to the corresponding sample label space. Its essence is to use several  $1 \times 1$  volumes. The product core performs convolution compression on the upper layer feature map, transforms the two-dimensional feature expression into a one-dimensional vector representation, and maps the original feature to each implicit semantic node.

### 3. Design of the Convolutional Neural Network Recognition Model

*3.1. Design of Activation Function.* One problem with the traditional activation function is the disappearance of the gradient. From the previous back propagation algorithm derivation process, it can be seen that when the error value is reverse gradient descent, it must be multiplied by the current input  $x$ -value. Taking the sigmoid function as an example,  $\text{Grad} = \text{Error} * \text{Sigmoid}'(x) * x$ . Both types of functions have the characteristic of double-ended “saturation”; that is, the value range is limited. If the derivative value is in the  $(0, 1)$  interval, the derivative value is scaled, and if the  $x$ -value is in  $(0, 1)$  or  $(-1, 1)$ , the saturation value is scaled. In this way, the error value in the propagation process of each layer is reduced at a double speed. After multiple recursive back propagations, the gradient continues to attenuate and disappear, making the process of network model learning gradually slow. This is the vanishing gradient problem caused by sigmoid-like functions.

At this stage, the most commonly used activation function in neural network models is ReLUs. This function was created in a restricted Bozeman machine and has been used extensively in neural network models. The ReLUs activation function is the parameter entity itself when the parameter is positive and has a value of 0 when the parameter is negative, and its positive gradient is 1, and only one end is saturated. Based on this feature, the ReLUs function has two main characteristics: one is that the neurons output from it are more sparse; the other is that it can reduce the problem of gradient disappearance to a certain extent.

However, ReLUs is a nonnegative function, and its output activation average value is greater than zero. The mean value of the neural unit is nonzero, and as the input of the next layer, there will be an offset difference. In this way, with the gradual accumulation, the higher the layer of neurons, the greater the offset difference. The effect of too large offset difference is to reduce the learning speed of the neural network model. Some solutions are to adjust the weight update during the gradient descent process, such as the SReLU's function.

In order to make the average output of neurons tend to 0 as much as possible, some variants of ReLUs activation function are produced. For example, the function Leaky ReLUs replaces the 0 value part of ReLUs with a linear function  $f(x) = ax$ , and other similar methods such as parameterized ReLUs and random leaky ReLUs approximate linear changes. Although the activation functions LReLU's, PReLU's, and RReLU's can make the average output of neurons 0, they still do not guarantee that neurons are in a noise-robust activation state.

*3.2. Design of Convolution and Pooling.* In the existing convolutional neural network model, the convolution operation is an indispensable operation, and there is no fixed requirement for the design of the convolution kernel size. In

general, the size of the convolution kernel is related to the “depth” design of the model. In the case of the same input size of the 3D visual image of aerobics, the larger the convolution kernel of each layer, the faster the size of the feature map will shrink. This paper proposes that the N-VGG network is based on the VGG network transformation, in which the fixed value of the convolution kernel is set to the size of  $3 * 3$ , and the moving step is 1. There are some feature maps at the edge of the  $3 * 3$  visual image of aerobics movements that cannot just meet the size of the  $3 * 3$  sliding window. At this time, a “patch” operation is required; that is, the last few pixels of the 3D visual image of aerobics movements are insufficient.

Existing convolutional neural network models generally use maximum pooling or average pooling. The difference between maximum pooling and mean pooling is the value of the function evaluation. There are also some pooling methods that are somewhat different in operation, such as overlapping pooling, which only sets the moving step of the sliding window to be smaller than the side length of the window. Therefore, adjacent pooling windows will have some overlapping areas, and the corresponding overlapping pooling generated features maps are also larger. Overlapping pooling can improve the recognition accuracy of CNN to a certain extent for the three-dimensional visual images of aerobics with tight final features. Here we introduce a new pooling method: Spatial Pyramid Pooling (SPP). The SPP pooling method utilizes the multiscale information of the pooled area and can solve the problem of different input sizes of the three-dimensional visual image of aerobics.

The SPP pooling method comes from a defect in the traditional CNN model; that is, the size of the input aerobics action 3D visual image in the traditional CNN model is fixed, and the 3D visual image of aerobics action of different sizes is usually scaled. The zooming operation will cause the distortion of the three-dimensional visual image of the aerobics action to a certain extent, which may lose some pixel spatial information, and the recognition accuracy of the model will be affected. SPP pooling can map different sizes of aerobics 3D visual images to the same feature dimension. Since SPP can process 3D visual images of aerobics movements of any size, it can also avoid the zooming operation of feature maps or original maps, which can reduce the information loss of 3D visual images of aerobics movements to a certain extent, thereby improving the 3D visual images of aerobics movements characteristic expression ability. The specific operation is relatively simple, as shown in Figure 3.

The feature map is pooled into three feature maps, corresponding to 1, 4, and 16 feature points. The middle connection layer contains 256 feature points. These 256 feature points and the previous 21 feature points are, respectively, fully connected, and after being fully connected, they are cascaded into a 5376-dimensional feature vector. We perform the next step on this feature vector. Spatial pyramid pooling is to map feature maps of different scales to feature vectors of the same dimension in this way.

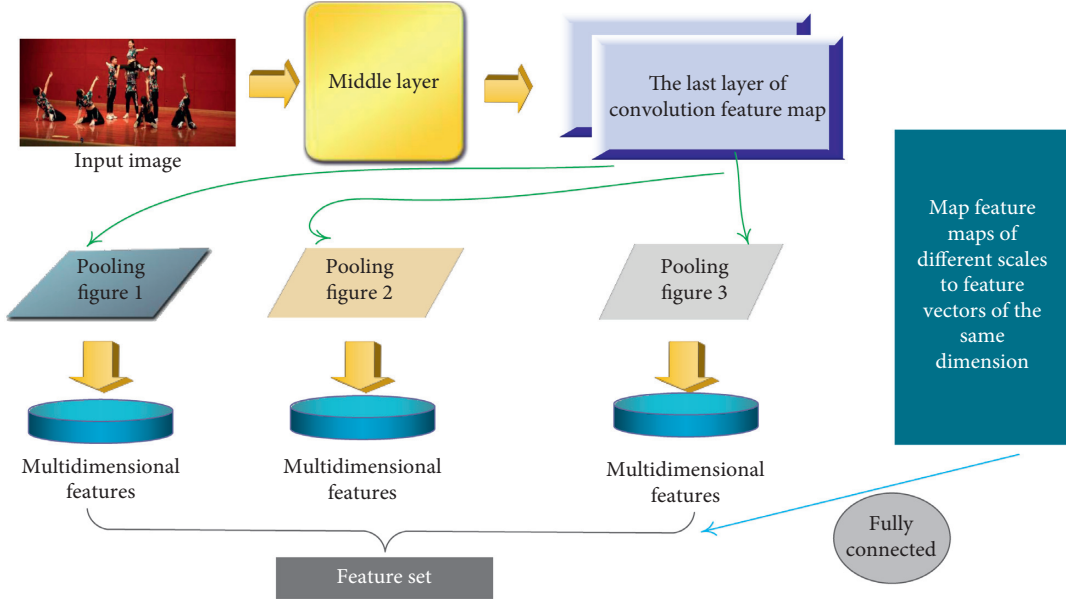


FIGURE 3: Schematic of spatial pyramid pooling.

3.3. *Recognizer Design.* Support vector machine is a very classic recognition algorithm in the field of machine learning. “Machine” in the field of machine learning usually refers to algorithms, and support vectors refer to variables that can affect decisions.

Suppose there are several training samples  $(x_i, y_i)$ , which can be separated by the hyperplane  $L3: wx + b = 0$ . At this time, the distance between  $L3$  and the closest sample point of the two categories is maximized. In the above linearly separable case, the parallel hyperplane line defining the support vector point is  $L1: w * x + b = 1$  and  $L2: w * x + b = -1$ , then the distance between the two is  $2/\|w\|$ ; at this time, the maximum identification interval is maximized; that is,  $\|w\|$  is minimized. At this time, all sample points are satisfied:

$$\begin{cases} x_i \cdot w \geq -b + 1, & y_i = 1, \\ x_i \cdot w < -b - 1, & y_i = -1. \end{cases} \quad (8)$$

At this point, the problem is equivalently defined as

$$\begin{aligned} & \min_{w,b} 0.5\|w\|^2, \\ & \text{s.t. } y_i(x_i \cdot w + b) > 1, \quad i = 1, 2, \dots, n-1, n. \end{aligned} \quad (9)$$

The above problem becomes a minimum value solving problem under several inequality constraints, that is, an optimization problem of convex quadratic programming. To construct the Lagrange function for the above problem, we get

$$L(w, a, b) = 0.5w^T \cdot w - \sum_{i=1}^n a_i \cdot y_i (x_i \cdot w + b - 1). \quad (10)$$

Among them,  $a_i$  is the Lagrange multiplier. According to the Lagrange duality theory, the dual problem of the original problem can be obtained. This problem can be solved by the SMO algorithm, and the solution of  $w$  can be obtained as

$$w = \sum_{i=0}^{n-1} a_i \cdot x_i \cdot y_i. \quad (11)$$

The decision function at this time is

$$f(x) = \text{sgn} \left[ \sum_{i=0}^{n-1} a_i \cdot y_i \cdot (x, x_i) + b \right]. \quad (12)$$

In this paper, the improved N-VGG convolutional neural network model based on the VGG network model is basically the same as the original VGG network model. The difference in the framework includes the introduction of the IRPN network at the last shared convolutional layer, which is used for model segmentation. The pooling layer after the convolutional layer is changed from the original maximum pooling layer to the spatial pyramid pooling layer.

In addition, inspired by the IRPN network, a bounding box regression layer is also introduced in the recognition layer. The function of the bounding box regression layer is the same as the bounding box regression method in IRPN. Its main function is to correct the bounding box of the target during the test phase. The bounding box regression at this time is the regression correction of the original target and the predicted target, so the loss function is

$$L(p, u, v, t^u) = L_{\text{loc}}(t^u, v) + L_{\text{cls}}(p, u). \quad (13)$$

The pooling of the SPP layer is performed on the cells after cutting. Considering that the pixel space of  $2 * 2$  cells is small enough, fully connected mapping is used for  $2 * 2$  cells. Here, only  $7 * 7$  and  $4 * 4$  cells are pooled in spatial pyramid. Among them, the size of the intermediate connection layer is changed to 179 dimensions. At this time, the dimension of the first fully connected layer is 4053.

## 4. Simulation Experiment and Result Analysis

*4.1. Convolutional Neural Network Aerobics 3D Visual Image Recognition Simulation.* By adjusting the configuration parameters of the convolutional neural network to improve the recognition accuracy of the network, we obtain relevant experimental data and analyze and discuss the ways and costs of improving the recognition results of the convolutional neural network.

Since the 3 data sets of the experiment are all grayscale images with small pixels, according to the principle of convolutional neural network, suppose the image size is  $n * n$ , the size of the first convolution kernel is  $c1 * c1$ , and the size of the downsampling area is  $s1 * s1$ ; the output feature map size after processing by the first convolutional layer and downsampling layer is  $o1 = (n + 1 - c1) / s1$ . The division operation also needs to be divided. Therefore, in order to ensure that the convolution kernel and sampling area have the appropriate size, the designed convolutional neural network has only 2 convolutional layers and 2 downsampling layers. A small number of network layers can also prevent the back-propagation algorithm from transmitting the error update value from back to front without becoming too small due to too many layers. The final fully connected layer will be adjusted for the number of categories of different data sets.

In the training method of the convolutional neural network, this experiment uses the BP algorithm for training; that is, first, we randomly initialize the convolution kernel matrix and then update the convolution kernel according to the error between the training result and the real result. Different from the BP algorithm, this experiment uses a method called min-batch for iteration. This method does not calculate the error of all training samples and update the weight parameters in each iteration; instead, the training samples are divided into multiple blocks. (Batch); each iteration only trains a sample of one block and adjusts the weight according to the error of the sample of the block. Since the weight update is performed for some samples, the learning rate can be increased appropriately to increase the update range of the weight on the partial samples, so that it can quickly adapt to all samples. Training the training samples in blocks helps to speed up the update speed of weights and improve training efficiency. The network structure (6, 5) – 2 – (12, 5) – 2 shows that the convolutional neural network has a total of 2 convolutional layers and 2 downsampling layers. There are 6 convolution kernels in the first convolutional layer. The size of each convolution kernel is  $5 * 5$ , and the size of the first downsampling area is  $2 * 2$ ; there are 12 convolution kernels in the second convolution layer, and the size of each convolution kernel is  $5 * 5$ . The size of the two downsampling areas is  $2 * 2$ .

Some parameters of the convolutional neural network are adjusted to obtain network configurations with different structures. Experiments were performed on these network models with different configurations to obtain experimental results. We conducted a total of 10 sets of experiments. Figure 4 shows the test results of the convolutional neural network on the aerobics dataset.

The convolutional neural network is trained in the min-batch method, and each block has a size of 50, which means 50 training samples are included. Before the experiment, the

training samples were randomly shuffled to ensure that the sample distribution in each block is random. The training data set has a total of 80,000 samples, and parameter updates are performed after each block is trained. Therefore, a total of  $80,000 / 50 = 1600$  parameter updates are required for one iteration. Figure 5 is the training error graph of experiment 1.

It can be seen from the error curve in Figure 5 that the min-batch method of training makes the error of the network model fluctuate, but all are within a certain range. Because the update parameters are performed after each small sample of training, the sample is not global, so it is not ruled out that there may be jitter when the error drops, but the error is still acceptable from the overall situation.

The network parameters used in experiment 1 and experiment 2 are exactly the same; the only difference is the sampling method. Experiment 1 uses mean sampling; experiment 2 uses maximum sampling. Experiment 2 is more advantageous in terms of efficiency. Because only the maximum value is taken, experiment 2 avoids the mean value calculation process. If the sampling area is large, the running time advantage of experiment 2 will be more obvious. And the maximum value sampling can well represent the difference of the data in the sampling area, reflect the most characteristic data in the area, reduce information loss, and improve the recognition accuracy. Therefore, in experiments 2 to 10, maximum sampling is used.

Experiments 2, 5, and 6 used 3, 6, and 12 convolution kernels in the first convolution layer when the size of the convolution kernels was the same. As the running time becomes longer, the experimental results are not gradually improved. The specific relationship is shown in Figure 6.

It can be seen from Figure 6 that when the size of the convolution kernel is constant, the recognition accuracy shows a wave trend as the number of convolution kernels increases. This shows that the increase in the number of convolution kernels is beneficial to the extraction of more features, thereby improving the recognition effect. However, when a certain number of convolution kernels is reached, the recognition effect is not the highest. The reason is that the extracted features are so many that redundant noise is also extracted, resulting in an overfitting phenomenon, resulting in a reduction in the recognition result.

When the number of the first layer of convolution kernels is certain, increasing the size of the convolution kernel can improve the recognition result because the increase of the convolution kernel can enlarge the area of the extracted features, so that the associated features in the area can be better extracted. However, a too large convolution kernel will also cause a decrease in results because the expansion of the extraction area leads to a loss of information. The comparative experiments 2, 3, and 4 are shown in Figure 7.

Comparing experiments 3, 7, and 8, the network structures used by them are exactly the same, but experiment 2 is iterated once during training; experiment 7 is iterated 5 times, a total of  $1600 * 5 = 8000$  parameter updates; and experiment 8 iterates 50 times, a total of  $1600 * 50 = 80000$  parameter updates. Figure 8 shows the relationship between training time and recognition accuracy. According to the curve, we can find that increasing the training time in the first part can

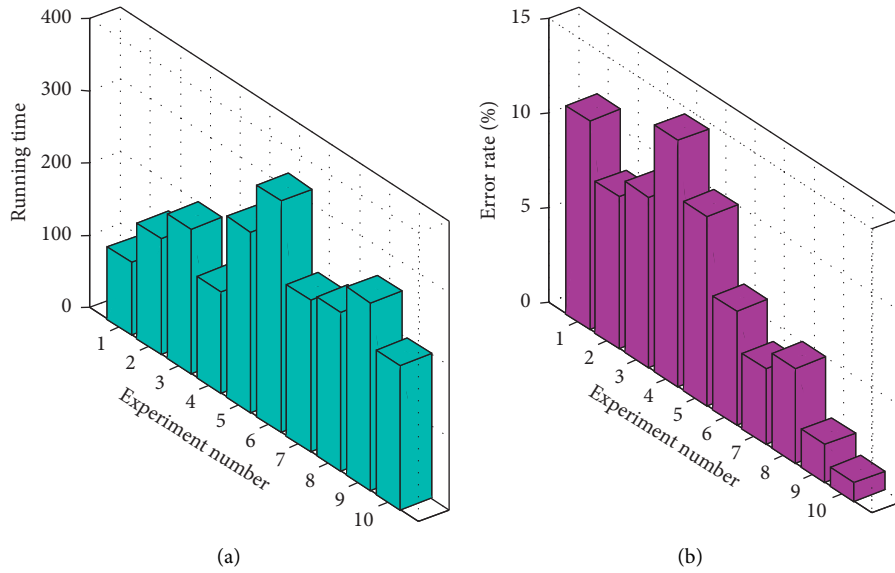


FIGURE 4: Test results of convolutional neural network on the data set. (a) Running time. (b) Error rate.

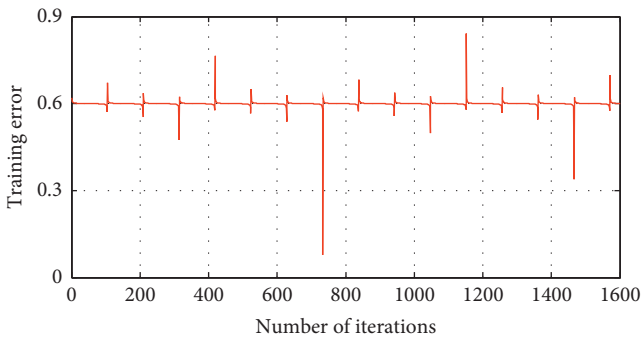


FIGURE 5: 3D visual image experiment of aerobics movement 1 training error.

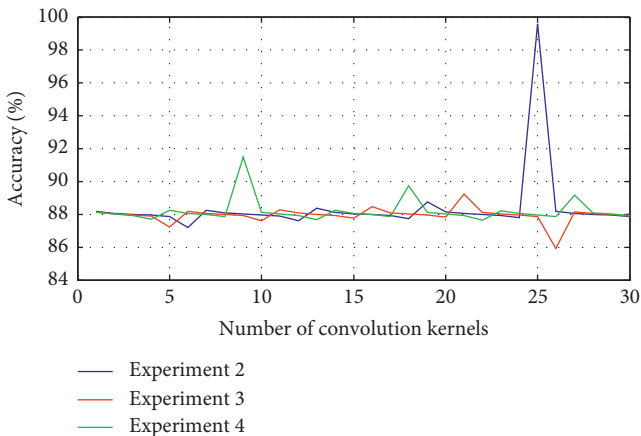


FIGURE 6: The relationship between the number of convolution kernels and the recognition results.

effectively improve the accuracy of the model, but the accuracy of the latter part will increase slowly or even decrease. Therefore, increasing the training time does not produce a

significant improvement in accuracy. Due to the limitations of hardware conditions, this experiment did not conduct a higher number of iterations, but it is theoretically expected that as the number of iterations increases, the network may overfit the training data set, thereby affecting the results on the test data set.

**4.2. Deep Neural Network Aerobics 3D Visual Image Recognition Simulation.** The deep learning neural network designed in this section is based on an autoencoder, stacking multiple autoencoders, and the hidden layer of the previous encoder is used as the input of the latter encoder, thus forming a multi-autoencoder deep neural network. Because the hidden layer of the autoencoder extracts the potential features of the data, using it as the input of the subsequent autoencoder can further extract the features to obtain more abstract high-level features. The results of the 3D visual image recognition of aerobics by the deep neural network are shown in Figure 9.

By adjusting the structure and configuration of the deep neural network, we got the following experimental results. Figure 10 shows the test results of the network model on the data set.

Increasing the number of hidden layers can also increase the final accuracy to a certain extent, but the accuracy will decrease when it reaches a certain number. The reason for the analysis is that increasing the number of hidden layers means increasing the number of abstractions of shallow features, which makes the later features more advanced, thus improving the recognition results. However, the autoencoder used in the experiment will minimize the error when extracting features of each layer, but this does not mean that no error will be generated. The increase in the number of layers will lead to the accumulation of errors, so that the error will be reduced to a certain extent. According to Figure 11, we can observe that when the number of features

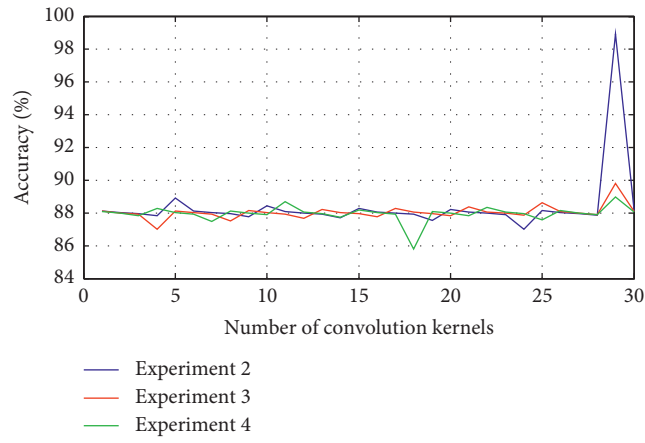


FIGURE 7: The relationship between the size of the convolution kernel and the recognition result.

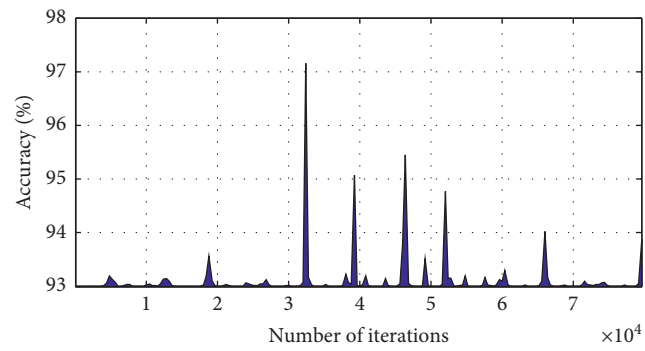


FIGURE 8: The relationship between training time and recognition accuracy.



FIGURE 9: Recognition result of 3D visual image of aerobics action by deep neural network.

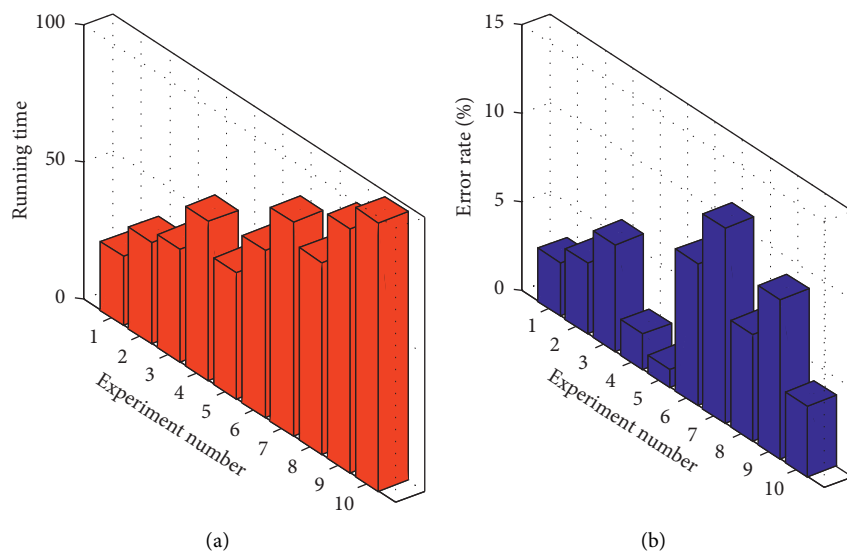


FIGURE 10: Test results of the deep neural network on the data set. (a) Running time. (b) Error rate.

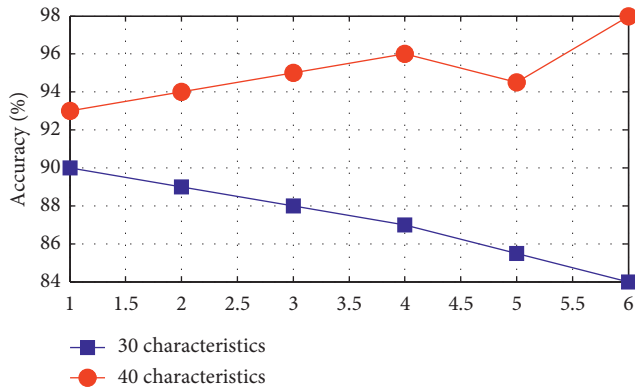


FIGURE 11: Comparison of the number of hidden layers and the number of nodes of the autoencoder.

is relatively large, the increase in the number of hidden layers can significantly improve the recognition accuracy.

## 5. Conclusion

This paper, through the research of aerobics action 3D visual image recognition and deep learning technology, introduces in detail the aerobics action 3D visual image recognition and deep learning theory, summarizes the description of common algorithms for feature extraction, and elaborates common aerobics such as R-CNN. The action 3D visual image recognition algorithm model and performance indicators verify the performance of the algorithm model. This paper introduces a new activation function ELU to replace the traditional functions such as ReLU. In addition, the maximum pooling layer of the last layer of the N-VGG model becomes spatial pyramid pooling. We train an SVM recognizer in the test phase to compare the recognition performance of the softmax recognizer. Convolutional neural network and autoencoder-based deep learning neural network are used as experimental objects, and they are applied to the benchmark data set for recognition testing. By adjusting the network configuration parameters and conducting multiple comparison experiments, the accuracy of the neural network model is improved. In the experiment, the model fusion method based on the improved AdaBoost algorithm is implemented and the above neural network model is fused. By comparing the experimental results of the model, we conclude that the model fusion method based on the improved AdaBoost algorithm is not accurate in the fusion model. In the case, when there is only a high accuracy model, there is also an improvement, but the effect is not as obvious as the first case.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this paper.

## References

- [1] J. Li, Y. Wang, J. See, and W. Liu, "Micro-expression recognition based on 3D flow convolutional neural network," *Pattern Analysis and Applications*, vol. 22, no. 4, pp. 1331–1339, 2019.
- [2] J. Cai and J. Hu, "3D RANs: 3D residual attention networks for action recognition," *The Visual Computer*, vol. 36, no. 6, pp. 1261–1270, 2020.
- [3] Z. Chen, H. Gao, Z. Zhang, H. Zhou, X. Wang, and Y. Tian, "Underwater salient object detection by combining 2D and 3D visual features," *Neurocomputing*, vol. 391, pp. 249–259, 2020.
- [4] R. D. Singh, A. Mittal, and R. K. Bhatia, "3D convolutional neural network for object recognition: a review," *Multimedia Tools and Applications*, vol. 78, no. 12, pp. 15951–15995, 2019.
- [5] J. Cao, Y. Hu, B. Yu, R. He, and Z. Sun, "3D aided duet GANs for multi-view face image synthesis," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 8, pp. 2028–2042, 2019.
- [6] G. Lin, Y. Zhang, G. Xu, and Q. Zhang, "Smoke detection on video sequences using 3D convolutional neural networks," *Fire Technology*, vol. 55, no. 5, pp. 1827–1847, 2019.
- [7] W. Sun, H. Zhao, and Z. Jin, "A facial expression recognition method based on ensemble of 3D convolutional neural networks," *Neural Computing and Applications*, vol. 31, no. 7, pp. 2795–2812, 2019.
- [8] Z. Liu, G. Song, J. Cai, T.-J. Cham, and J. Zhang, "Conditional adversarial synthesis of 3D facial action units," *Neurocomputing*, vol. 355, pp. 200–208, 2019.
- [9] N. Hättenschwiler, M. Mendes, and A. Schwaninger, "Detecting bombs in X-ray images of hold baggage: 2D versus 3D imaging," *Human Factors*, vol. 61, no. 2, pp. 305–321, 2019.
- [10] R. Mehrizi, X. Peng, X. Xu, S. Zhang, and K. Li, "A deep neural network-based method for estimation of 3D lifting motions," *Journal of Biomechanics*, vol. 84, pp. 87–93, 2019.
- [11] T. Akilan, Q. J. Wu, A. Safaei, H. Jie, and Y. Yimin, "A 3D CNN-LSTM-based image-to-image foreground segmentation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 959–971, 2019.
- [12] M. N. I. Qureshi, J. Oh, and B. Lee, "3D-CNN based discrimination of schizophrenia using resting-state fMRI," *Artificial Intelligence in Medicine*, vol. 98, pp. 10–17, 2019.
- [13] J. Nie, N. Xu, M. Zhou, G. Yan, and Z. Wei, "3D Model classification based on few-shot learning," *Neurocomputing*, vol. 398, pp. 539–546, 2020.
- [14] Z. Zhu, C. Rao, S. Bai, and L. J. Latecki, "Training convolutional neural network from multi-domain contour images for 3D shape retrieval," *Pattern Recognition Letters*, vol. 119, pp. 41–48, 2019.
- [15] X. Lu, H. Yao, S. Zhao, X. Sun, and S. Zhang, "Action recognition with multi-scale trajectory-pooled 3D convolutional descriptors," *Multimedia Tools and Applications*, vol. 78, no. 1, pp. 507–523, 2019.
- [16] C. L. Chowdhary, "3D object recognition system based on local shape descriptors and depth data analysis," *Recent Patents on Computer Science*, vol. 12, no. 1, pp. 18–24, 2019.
- [17] R. Singh, J. K. Dhillon, A. K. S. Kushwaha, and R. Srivastava, "Depth based enlarged temporal dimension of 3D deep convolutional network for activity recognition," *Multimedia Tools and Applications*, vol. 78, no. 21, pp. 30599–30614, 2019.
- [18] L. E. Carvalho and A. von Wangenheim, "3D object recognition and classification: a systematic literature review,"

*Pattern Analysis and Applications*, vol. 22, no. 4, pp. 1243–1292, 2019.

- [19] F. Fiorucci, F. Ardizzone, A. C. Mondini, A. Viero, and F. Guzzetti, “Visual interpretation of stereoscopic NDVI satellite images to map rainfall-induced landslides,” *Landslides*, vol. 16, no. 1, pp. 165–174, 2019.
- [20] P. Wei, H. Sun, and N. Zheng, “Learning composite latent structures for 3D human action representation and recognition,” *IEEE Transactions on Multimedia*, vol. 21, no. 9, pp. 2195–2208, 2019.



## Research Article

# Moving Vehicle Tracking Optimization Method Based on SPF

Caixia Lv<sup>1,2</sup> and Xuejing Zhang<sup>2</sup>

<sup>1</sup>School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China

<sup>2</sup>Smart City College, Beijing Union University, Beijing 100101, China

Correspondence should be addressed to Caixia Lv; [xxtcaixia@bnu.edu.cn](mailto:xxtcaixia@bnu.edu.cn)

Received 16 September 2020; Revised 19 October 2020; Accepted 24 October 2020; Published 12 November 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Caixia Lv and Xuejing Zhang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the intelligent transportation system, the license information can be automatically recognized by the computer and the vehicle can be tracked. Red light running, illegal change of lanes, vehicle retrograde, and other illegal driving events are reasonably recorded. This is undoubtedly an effective help for the traffic police to relieve the huge work pressure. However, in China, a considerable number of vehicle tracking methods have certain limitations in resisting complex external environmental influences. The external environmental factors include but not limited to variable factors such as camera movement, jitter, and severe rain and snow. These factors cannot be controlled well, so the tracking accuracy is greatly reduced. In regard to this, this paper proposes an optimization method for moving vehicle tracking based on SPF. First, according to the size of the overlapping area of the motion area between the two images, the researcher can construct and simplify the vertex adjacency matrix that reflects the characteristics of the undirected bipartite graph. Then according to the corresponding relationship between the vertex adjacency matrix and the regional behavior and vehicle behavior, the researcher completes the regional behavior analysis and vehicle behavior analysis. On this basis, a particle filter vehicle tracking algorithm based on segmentation compensation is introduced, and the vector sum of the tracked segmentation area is used as the final position of the target vehicle. In this way, as many scattered particles fall on the target area as possible, which will greatly improve the efficiency of particle utilization, enhance tracking accuracy, and avoid the problem of tracking failure caused by too fast vehicle movement. Through experimental simulation, it can be seen that the method proposed in this paper can greatly enhance the vehicle tracking ability when tracking vehicles in “complex environments.”

## 1. Introduction

In the application of target tracking that involves getting close to people's livelihood, one of the most critical areas is vehicle tracking. Vehicles on the roads in China have gradually transitioned from being dominated by bicycles to being dominated by automobiles with the continuous improvement of related technologies in the industry and the great optimization of residents' quality of life. Family ownership of cars is about to leap forward, and more and more families own their own cars, making road traffic more and more burdensome. On the other hand, urban construction has been finalized, and roads in the city have nowhere to be repaired. The limited number of roads corresponds to the increasing number of cars, causing frequent traffic accidents. These problems put forward more stringent

requirements on road traffic management and gradually formed the concept of an intelligent transportation system. Intelligent transportation can use the sensitivity of the computer to identify the information of passing vehicles and record the route of the vehicle in real time. In this way, any traffic violations such as running a red light, illegally changing lanes, and vehicles going backward will be recorded. It will greatly reduce the work pressure of traffic police, improve driving safety, and reduce traffic accidents.

Academic research in the field of vehicle tracking is very complicated. With the rapid development of information technology, video surveillance has penetrated into all areas of people's production and life and naturally attracted the attention of scholars from all over the world. There are many countries that have invested a lot of manpower, material, and financial resources to in-depth research and solved problems

in the field of vehicle tracking, which better promotes the rapid development of the transportation field. The 8<sup>th</sup> literature introduced the Kalman filter algorithm and related mathematical theories into the target tracking algorithm, which promoted the development of vehicle tracking technology during this period. However, this method is difficult to adapt to the weather, environment, and external light factors, which will lead to the problem of low vehicle tracking accuracy. The 9<sup>th</sup> literature made outstanding contributions in several aspects such as mean shift theory, probabilistic multihypothesis tracking, variable structure multimodel, and distributed target fusion, which had pushed the field of vehicle tracking forward. But this method was subject to fall into the optimal solution. The 10<sup>th</sup> literature proposed a vehicle tracking method based on feature extraction of vehicle edges. In this method, the target edge information was obtained by detecting the changing gray value, then the vehicle characteristics were proposed, and the tracking of the vehicle was completed on this basis. This method had high stability to external factors such as light. But there is a problem that tracking takes a long time.

For this reason, an SPF-based optimization method for moving vehicle tracking is proposed in this paper. Particularly for the background failure due to the influence of weather interference, changing the optical fiber and other dynamic factors on the background, according to the size of the overlapping area of the moving area between the two frames of images, this paper constructs and simplifies the vertex adjacency matrix reflecting the characteristics of the undirected bipartite graph. Then, according to the corresponding relationship between the vertex adjacency matrix and the regional behavior and vehicle behavior, a particle filter algorithm (SPF) based on segmentation compensation is proposed for the problem that the vehicle tracking model is easy to fall into the optimal solution [1]. During the process of tracking the target vehicle, this algorithm divides the target into multiple regions. Based on the local features of the selected target in these regions, if the tracked target is partially occluded or the shape changes, other regions of the target can be used instead of tracking, which can make the particle filter more accurate and applicable. Regarding the time-consuming problem of the vehicle tracking process, this paper adds the resampling calculation after calculating the weight of the particles. The particles are resampled according to the weight of the particles, and particles with small weights are deleted, which reduces the amount of calculation time.

## 2. Behavior Analysis of Moving Vehicles Based on Directionless Two-Part Graph

**2.1. Principles of Vehicle Behavior Analysis.** The directionless two-part graph is a very useful mathematical model, which has important applications in many scientific fields. The directionless two-part graph can be used to describe many problems, for example, “data allocation,” “work arrangement,” “time allocation,” and “personnel selection.” This paper combines the directionless two-part graph with the behavior analysis of the vehicle in the video and uses the

vertex adjacency matrix representing the directionless two-part graph to determine the behavior of the moving area. Besides, by combining the relationship between regional behavior and vehicle behavior, this paper completes vehicle behavior analysis. The overall flowchart of the algorithm is shown in Figure 1.

The biggest advantage of applying the directionless two-part graph to vehicle behavior analysis is that the method does not require high accuracy for moving target extraction. Even if the extracted moving regions are not continuous, vehicle behavior analysis can be performed.

**2.2. Vehicle Analysis Model Based on Directionless Two-Part Graph.** The directionless two-part graph is expressed by the vertex adjacency matrix, the essential meaning and function of the two are the same, but the vertex adjacency matrix can better reflect the association relationship of each area and is more suitable for programming. The following paper will introduce the basic principles and implementation of vertex adjacency matrix construction. The flowchart of vertex adjacency matrix construction can be seen in Figure 2.

On the basis of extracting the moving vehicle area, the researcher can transform the behavior analysis of the vehicle into the behavior analysis of the area. The behavior analysis of the region is obtained by analyzing the vertex adjacency matrix constructed below.

This paper extracts the moving area in each frame of the traffic surveillance video and replaces the original moving area with the smallest bounding rectangle of each area [2]. The change of the motion area between two adjacent frames of images can be defined as four behaviors: appearance, disappearance, merging, and decomposition, as shown in Figure 3. Figure 3(a) describes the three motion regions extracted in the frame as R1, R2, and R3; Figure 3(b) describes the three motion regions extracted in the first frame, denoted as R4, R5, and R6. From Figures 3(a) and 3(b), the relationship between the motion areas in two adjacent frames can be obtained intuitively, as shown in Figure 3(c). In Figure 3(a), part of R1 is decomposed into R5, and the other part is merged with R2 to form R4. R3 disappears because there is no associated motion area, and R6 is the newly appeared area in Figure 3(b).

The directionless two-part graph can describe the relationship between the motion areas in adjacent frames. Within the frame  $k - 1$  and  $k$  in the image, this paper takes the upper left vertex of each motion area as the vertex sets  $U = \{U_1, U_2, \dots, U_m\}$  and  $V = \{V_1, V_2, \dots, V_n\}$ , respectively. Each vertex represents a motion area; that is, the vertex set  $U = \{U_1, U_2, \dots, U_m\}$  corresponds to the motion area set  $Ru = \{Ru_1, Ru_2, \dots, Ru_m\}$  in the frame  $k - 1$ ; the vertex set  $V = \{V_1, V_2, \dots, V_n\}$  corresponds to the motion area set  $Rv = \{Rv_1, Rv_2, \dots, Rv_n\}$  in the frame  $k$ .

Each edge represents the association relationship ( $i = 1, 2, \dots, m; j = 1, 2, \dots, n$ ) between the motion regions  $Ru_i$  and  $Rv_j$  of adjacent frames, and  $E$  is the set of all edges in  $G$ . The vertex adjacency matrix  $A_{n \times m}$ , which is equivalent to the directionless two-part graph  $G(U, V, E)$ , can clearly describe the edge set  $E$ . In vertex adjacency matrix  $A_{n \times m}$ , row  $m$

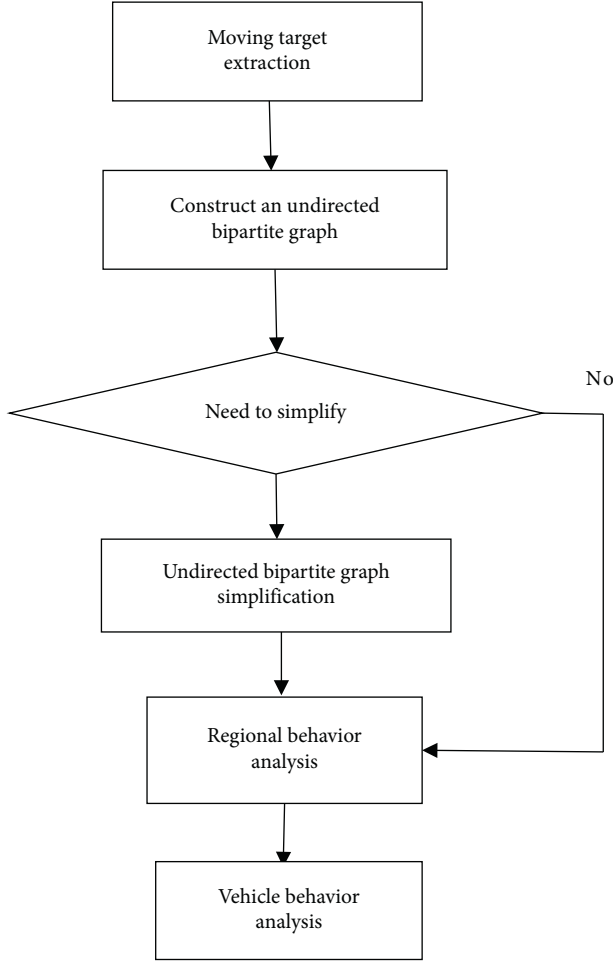


FIGURE 1: Flowchart of vehicle behavior analysis.

correspond to  $m$  vertices in vertex set  $U = \{U_1, U_2, \dots, U_m\}$ , that is, the  $m$  motion regions contained in the frame  $k-1$ ; the  $n$  columns correspond to the  $n$  vertices in the vertex set  $V = \{V_1, V_2, \dots, V_n\}$ , that is, the  $n$  motion regions contained in the frame  $k$ . Then the meaning of the element  $a_{ij}$  of the vertex adjacency matrix can be expressed as follows:

If  $a_{ij} = 1$ , it means that there is an edge between the vertices  $U_i$  and  $V_j$ ; that is, there is a correlation between the motion areas  $Ru_i$  and  $Rv_j$ .

If  $a_{ij} = 0$ , it means that there is no edge between the vertices  $U_i$  and  $V_j$ ; that is, there is no correlation between the motion regions  $Ru_i$  and  $Rv_j$ .

It can be seen from the above definition that whether or not there is a correlation between each motion area in the two images before and after determines the value of the vertex adjacency matrix element  $a_{ij}$ . The literature gives a method to determine the relationship between motion areas; that is, it is determined by calculating the overlap area between the motion areas  $Ru_i$  and  $Rv_j$ . The specific algorithm is shown in formula (1).

$$S(u_i, v_j) = \frac{\min(S(u_i), S(v_j))}{2} \quad (1)$$

In formula (1),  $S(u_i)$  represents the area of the motion area  $Ru_i$  in the frame  $k-1$ ,  $S(v_j)$  represents the area of the motion area  $Rv_j$  in the frame  $k$ ,  $S(u_i, v_j)$  represents the overlapping area of  $Ru_i$  and  $Rv_j$  in the motion area, and the relationship between  $S(u_i, v_j)$ ,  $S(u_i)$ , and  $S(v_j)$  is shown in Figure 4.

If equation (1) is satisfied between the movement areas  $Ru_i$  and  $Rv_j$ ,  $Ru_i$  and  $Rv_j$  are considered to be related,  $a_{ij} = 1$ .

If equation (1) is not satisfied between the movement areas  $Ru_i$  and  $Rv_j$ , then  $Ru_i$  and  $Rv_j$  are considered unrelated,  $a_{ij} = 0$ .

After constructing the vertex adjacency matrix according to the above algorithm, the behavior of the moving area can be analyzed through the simplification of the vertex adjacency matrix, and then the behavior analysis of the vehicle can be completed according to the conversion relationship between the area behavior and the vehicle behavior.

### 2.3. Optimization Analysis of Vehicle Behavior Based on Directionless Two-Part Graph

**2.3.1. Improvement of Directionless Two-Part Graph.** Theoretically, four behaviors such as merging, decomposing, appearing, and disappearing may occur in the motion area of adjacent frames. However, in a video stream with a sampling frequency of 25–30 frames/sec, the probability of simultaneous decomposition and merging of the same motion area is very small. This actual experience is reflected in the region relationship graph and vertex adjacency matrix and can be expressed as follows: the degree of the two vertices connected by each edge of  $G$  (i.e., the number of edges derived from each vertex) is at most one greater than or equal to 2. The sum of the number of “1” in the row and column where element  $a_{ij}$  is located cannot be greater than or equal to 2. So, in this paper, the edge with the degree of both vertices greater than or equal to 2 is called the interference edge, which will cause errors in regional behavior analysis. Therefore, before the regional behavior analysis, these interference edges must be removed by simplification of the directionless two-part graph.

In the actual surveillance video, there are noise interference and changes in the external environment. These factors will inevitably cause interference edges in  $G$ . And there may be multiple interference edges connected to a vertex, as for which one to delete, researchers need to follow the principle of simplification of the directionless two-part graph described below [3].

The improved logic of directionless two-part graph can be expressed as follows. First, by scanning the vertex adjacency matrix  $A_{m \times n}$ , the researcher records all interference edges, assuming that there are interference edges in  $M$  vertices. Then, for a certain vertex  $a_{ij}$  that has  $N$  interference edges connected to it, the researcher deletes  $l$  ( $l = 1, 2, \dots, N-1$ ) interference edges in turn to obtain  $C_N^1 C_N^2, \dots, C_N^{N-1}$  different subgraphs. In order to find the subgraph that is most consistent with the actual situation and define the matching function  $C$ , see the following

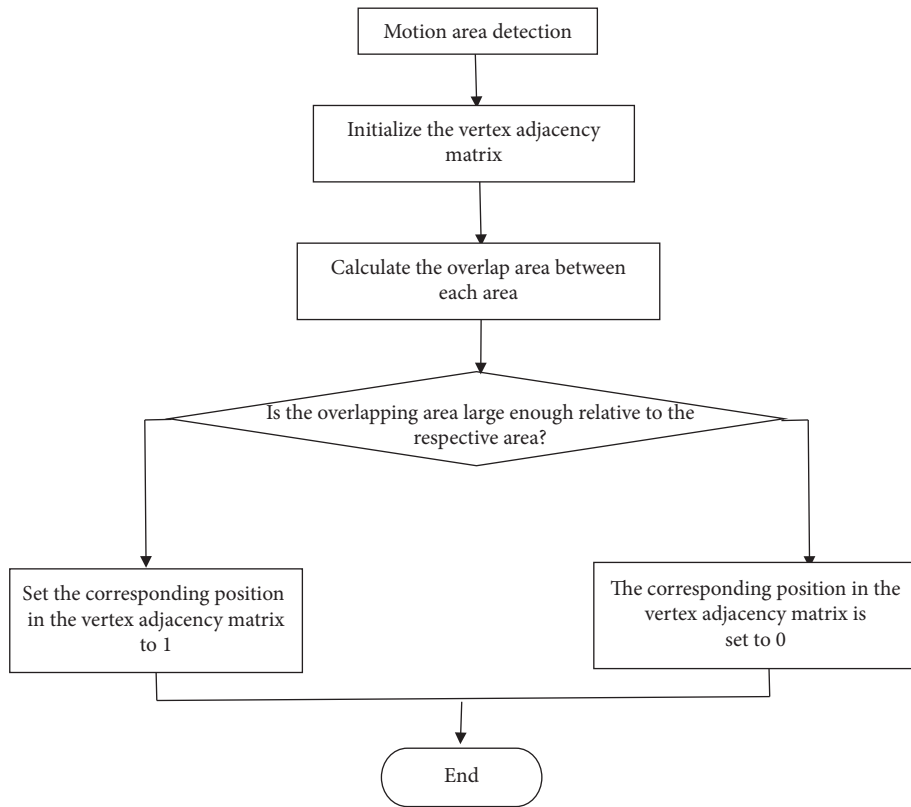


FIGURE 2: Flowchart of the construction of the vertex adjacency matrix.

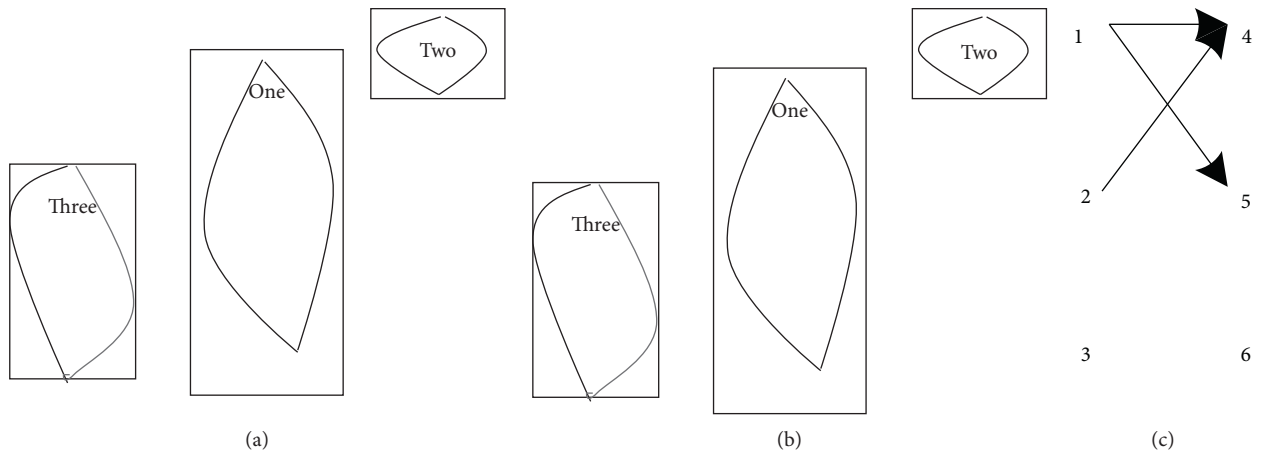


FIGURE 3: Four kinds of transformation in the region. (a) The area in frame  $k - 1$ . (b). The area in frame  $K$  (c). Correspondence between regions.

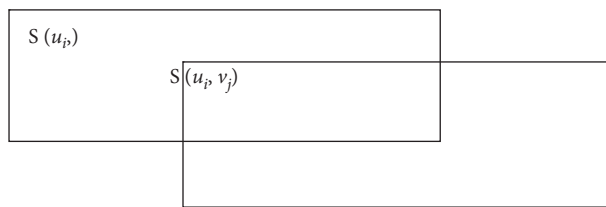


FIGURE 4: Calculation of the regions' relation.

formula (2); the researcher takes the subgraph with the smallest function value as the result of the vertex  $a_{ij}$ ; that is, deletes its interference edges according to the subgraph with the smallest function value and simultaneously updates the element values in the rows and columns of  $a_{ij}$  in the vertex adjacency matrix. In the same way, other vertices  $M - 1$  with interference edges also delete their interference edges and update the vertex adjacency matrix according to the same algorithm. When all  $M$  vertices are processed, the behavior of the motion area can be obtained according to the vertex adjacency matrix obtained by the simplification.

$$C(G) = \sum_{u \in P_k} \frac{|S(u_i) - A(u_i)|}{\max(A(u_i), S(u_i))}. \quad (2)$$

As for the directionless two-part graph, first, all vertices can be divided into “father vertices” and “subvertices.” The “father vertices” include all vertices with degree greater than 1 and degree equal to 0 in frame  $k - 1$  (the previous frame), and other vertices are “subvertices.” Assuming that the set of “father vertices” is  $P_k$ ,  $S(u_i)$  represents the area of the region  $Ru_i$  represented by the vertex  $U_i$  in the set  $P_k$ , and  $A(u_i)$  represents the sum of the areas of the motion regions represented by all vertices connected to the vertex  $U_i$ . Calculate the  $C(G)$  value of each subpicture, and take the subpicture with the smallest function value to describe the behavior of the motion area of the two adjacent frames [4].

Based on the above basic principles, the flowchart of the algorithm is shown in Figure 5.

**2.3.2. Vehicle Behavior Analysis.** Because there is a corresponding relationship between the vehicle and the moving area, the behavior analysis of the vehicle can be transformed into the behavior analysis of the moving area. However, there is not a simple one-to-one relationship between sports areas and vehicles. One car can correspond to multiple sports areas, and the same area may be shared by multiple vehicles. Table 1 summarizes the relationship between the features of the vertex adjacency matrix and the regional behavior and vehicle behavior.

From the vertex adjacency matrix, the logic of regional behavior and vehicle behavior can be expressed as follows. In the image of frame  $k - 1$ , assuming that the corresponding relationship between the vehicle and the moving region  $Ru_i$  is known, the final vertex obtained by simplification in the adjacency matrix  $A_{n \times m}$ , look for elements that meet the conditions in the first column of Table 1, then follow the rules in Table 1 to obtain the region behavior from frame  $k - 1$  to frame  $k$ , and then summarize the two-vehicle behavior between frames [5].

### 3. SPF-Based Vehicle Optimization Tracking

In the process of tracking a moving vehicle, because the vehicle is in constant motion, in addition to analyzing the behavior of the vehicle, it is also necessary to track it in real time. On the basis of understanding the relationship between the vertex adjacency matrix and the regional behavior and vehicle behavior, if the target is occluded during the

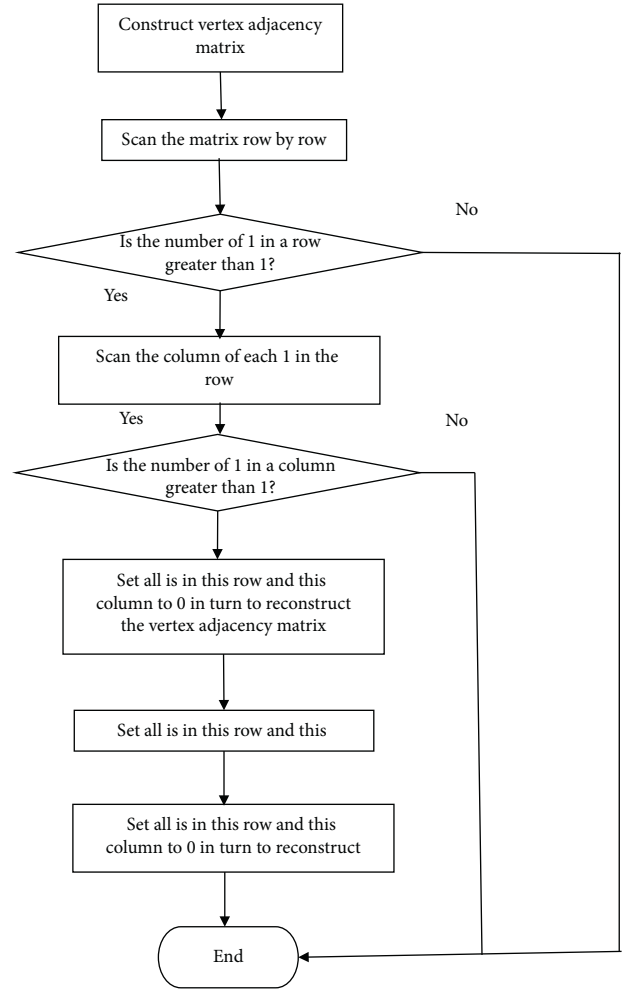


FIGURE 5: Flowchart of the simplification of the directionless two-part graph.

movement of the vehicle, it is necessary to grasp some unique characteristics of the vehicle itself. Tracking these features will achieve better tracking results.

#### 3.1. Moving Vehicle Tracking Based on SPF

**3.1.1. Principle of Particle Filter Algorithm.** As far as SPF and IPPF are concerned, they mainly extract effective tracking schemes by improving the correlation distribution on the standard particle filter. MC-JPDAF is the particle filter technology applied to the standard JPDAF, so as to effectively overcome the dimensionality problem by dividing the state space [6].

Based on Monte Carlo methods, the concept of standard particle filtering is proposed. It mainly uses particle sets to represent probability, which can be used in any form of state-space model. Its core idea is to express its distribution through random state particles drawn from the posterior probability, which is a kind of sequential importance sampling. To put it simply, the particle filter method refers to the process of approximating the probability density function by finding a set of random samples propagating in the

TABLE 1: Relation between the vertex adjacency matrix and the behaviors of region and vehicle.

| Vertex adjacency matrix features                 | Regional behavior                           | Vehicle behavior   |
|--|---|--|
| $a_{ij} = 1$ , all $a_{ih} \neq 1 (h \neq j)$    | $Ru_i$ and $Rv_j$ one-to-one correspondence | Drive normally   |
| $a_{ij} = 1$ , exists $a_{hj} \neq 1 (h \neq i)$ | $Ru_i$ and $Ru_h$ merge into $Rv_j$         | Obstructed entry (area in frame $k$ , both $Ru_i$ and $Ru_h$ are single vehicles)  |
| $a_{ij} = 1$ , exists $a_{ih} = 1 (h > j)$       | Decompose $Ru_i$ into $Rv_j$ and $Rv_h$     | Split (area $Ru_i$ in frame $k$ is a single vehicle)<br>Occlusion separation (occlusion exists in area $Ru_i$ in the frame $k$ ) |

state space and replacing the integral operation with the sample mean to obtain the state minimum variance distribution. The sample here refers to the particle. When the number of samples is  $N \rightarrow \infty$ , any probability density distribution can be approximated.

The basic principle of Monte Carlo method is as follows.

In a high-dimensional space, it is required to solve the following integral:

$$I(f) = \int_x f(X)p(X)dX. \quad (3)$$

In the above formula,  $p(X)$  is a probability distribution,  $f(\cdot)$  is any integrable function of  $p(X)$  and satisfies  $f: X \rightarrow R^n$ ,  $N$  independently distributed random sample functions  $p(X)$  are drawn from the probability distribution  $\{X^{(i)}\}_{i=1}^N$ , and then the probability distribution function  $p(X)$  can be approximated by the sample functions of these independent distributions as

$$p_N(X) = \frac{1}{N} \sum_{i=1}^N \delta(X - X^{(i)}). \quad (4)$$

Among them,  $\delta(X - X^{(i)})$  is the Dirac-Delta function, so the integral formula to be solved can become the following summation problem:

$$I_N(f) = \frac{1}{N} \sum_{i=1}^N f(X^{(i)}) \xrightarrow{N \rightarrow \infty} I(f) = \int_x f(X)p(X)dX. \quad (5)$$

This transformation from integration to addition is the Monte Carlo method. It is not difficult to conclude from the theorem of large numbers that  $I_N(f)$  is unbiased, and the Monte Carlo method converges to  $I(f)$ .

*Definition.*  $\sigma_f^2 \triangleq E_p(X)(f_2(X)) - I^2(f)$ , so the variance of  $I_N(f)$  is  $(\sigma_f^2/N)$ , and the following formula can be obtained from the central limit theorem:

$$\sqrt{N}(I_N(f) - I(f)) \xrightarrow{N \rightarrow \infty} N(O, \sigma_f^2). \quad (6)$$

If the probability distribution function  $p(X)$  obeys the Gaussian distribution, the random sample  $p(X)$  can be obtained directly through the sampling of  $\{X^{(i)}\}_{i=1}^N$ . However, in practical applications, in many cases  $p(X)$  does not satisfy the Gaussian distribution, so the Monte Carlo method can be used to solve this problem.

For discrete-time stochastic systems, in the known prior state distribution  $p(X_0)$ , the solution of  $p(X_{0:k}|Z_{1:k})$  can be obtained according to the Bayesian estimation method:

$$p(X_{0:k}|Z_{1:k}) = \frac{p(Z_k|X_k)p(X_k|X_{k-1})}{p(Z_k|Z_{1:k-1})} p(X_{0:k-1}|Z_{1:k-1}). \quad (7)$$

The posterior distribution  $p(X_{0:k}|Z_{1:k})$  can be solved by Monte Carlo method. In the case that the random process does not satisfy the Gaussian distribution, the mathematical expression of the posterior distribution is difficult to give intuitively, so it is difficult to directly obtain the posterior distribution. Defining  $q(X_{0:k}|Z_{1:k})$  is a conditional probability distribution that satisfies the non-Gaussian state, and it has a larger support set than the posterior distribution. Select  $N$  random samples from the recommended distribution to get a sample set  $\{X_{0:k}^{(i)}\}_{i=1}^N$ , which can get the following formula:

$$q(X_{0:k}|Z_{1:k}) = \frac{1}{N} \sum_{i=1}^N \delta(X_{0:k} - X_{0:k}^{(i)}). \quad (8)$$

Similarly, the expression of the posterior distribution can be written as

$$p(X_{0:k}|Z_{1:k}) = \frac{p(X_{0:k}|Z_{1:k})}{q(X_{0:k}|Z_{1:k})} q(X_{0:k}|Z_{1:k}) = \sum_{i=1}^N w_k^{(i)} \delta(X_{0:k} - X_{0:k}^{(i)}). \quad (9)$$

Among them,  $w_k^{(i)} = (\tilde{w}_k^{(i)} / \sum_{i=1}^N \tilde{w}_k^{(i)})$  is the normalized sample, which meets

$$\tilde{w}_k^{(i)} \propto \frac{p(X_{0:k}^{(i)}|Z_{1:k})}{q(X_{0:k}^{(i)}|Z_{1:k})}. \quad (10)$$

From the above mathematical reasoning, it can be seen that, in the non-Gaussian case, the posterior probability  $p(X_{0:k}^{(i)}|Z_{1:k})$  can be approximated by a set of random samples  $\{X_{0:k}^{(i)}, W_k^{(i)}\}_{i=1}^N$ . These random samples are called particles. The method of obtaining the posterior probability distribution through particle sampling is called importance sampling method.

When calculating the Bayesian state estimation, the above proposal distribution can be converted into a recursive form, with the following formula:

$$q(X_{0:k}|Z_{1:k}) = q(X_k|X_{0:k-1}, Z_{1:k})q(X_{0:k-1}|Z_{1:k-1}). \quad (11)$$

According to the Bayesian state estimation equation, the following formula can be obtained:

$$p(X_{0:k}|Z_{1:k}) \propto p(Z_k|X_k)p(X_k|X_{k-1})p(X_{0:k-1}|Z_{1:k-1}). \quad (12)$$

Put these two expressions into equation (12), it is not difficult to get

$$\begin{aligned} \tilde{w}_k^{(i)} &\propto \frac{p(Z_k^{(i)}|X_k^{(i)})p(X_k^{(i)}|X_{k-1}^{(i)})p(X_{0:k-1}^{(i)}|Z_{1:k-1})}{q(X_k^{(i)}|X_{0:k-1}, Z_{1:k})q(X_{0:k-1}^{(i)}|Z_{1:k-1})} \\ &= \tilde{w}_k^{(i)} \frac{p(Z_k^{(i)}|X_k^{(i)})p(X_k^{(i)}|X_{k-1}^{(i)})}{q(X_k^{(i)}|X_{0:k-1}, Z_{1:k})}. \end{aligned} \quad (13)$$

Adding a normalization factor to formula (13) can normalize  $\tilde{w}_k^{(i)}$ , so that, for the recursive Bayesian filtering problem,  $\tilde{w}_k^{(i)}$  can be rewritten as

$$\tilde{w}_k^{(i)} \propto \tilde{w}_{k-1}^{(i)} \frac{p(Z_k^{(i)}|X_k^{(i)})p(X_k^{(i)}|X_{k-1}^{(i)})}{q(X_k^{(i)}|X_{k-1}, Z_k)}. \quad (14)$$

In this way, the posterior probability  $p(X_k|Z_{1:k})$  can be approximately written as

$$p(X_k|Z_{1:k}) = \sum_{i=1}^N w_k^{(i)} \delta(X_k - X_k^{(i)}). \quad (15)$$

At this time, if the proposed distribution function is a state prior distribution, it can be derived as

$$q(X_k^{(i)}|X_{k-1}, Z_k) = p(X_k^{(i)}|X_{k-1}). \quad (16)$$

Bring the above formula into (15) is

$$\tilde{w}_k^{(i)} \propto \tilde{w}_{k-1}^{(i)} p(Z_k^{(i)}|X_k^{(i)}). \quad (17)$$

The process described above is the process of particle filtering. The entire process of particle filtering can be described as follows:

- (1) Initialization: at  $k = 0$ , according to the prior distribution  $p(X_0)$ , the initial particle set  $\{X_0^{(i)}, w_0^{(i)}\}_{i=1}^N$  is obtained, where  $w_0^{(i)} = (1/N)$ .
- (2) Sampling: for  $k = 1, 2, \dots$ , sample from the state transition probability relationship  $p(X_k^{(i)}|X_{k-1}^{(i)})$  to calculate a new particle set  $\{\tilde{X}_k^{(i)}\}_{i=1}^N$ .
- (3) Calculate the weight: calculate the weight  $\tilde{w}_k^{(i)}$  of the particle  $\tilde{X}_k^{(i)}$  according to  $\tilde{w}_k^{(i)} \propto \tilde{w}_{k-1}^{(i)} p(Z_k^{(i)}|X_k^{(i)})$ , and normalize the weight to obtain a new weight  $\{w_k^{(i)}\}_{i=1}^N$ .
- (4) Calculate the estimated state: according to the particle set  $p(X_k|Z_{1:k}) = \sum_{i=1}^N w_k^{(i)} \delta(X_k - X_k^{(i)})$  obtained in (17), the estimated state posterior

distribution can be calculated, and the state estimate is

$$\hat{X}_k = E(X_k) = \sum_{i=1}^N w_k^{(i)} \tilde{X}_k^{(i)}. \quad (18)$$

- (5) Resampling: according to the particle weight  $w_k^{(i)}$ ,  $N$  particles  $\{X_k^{(i)}\}_{i=1}^N$  are retrieved from the particle set  $\{\tilde{X}_k^{(i)}, w_k^{(i)}\}_{i=1}^N$ , and the new particle weight is  $(1/N)$ , thereby establishing a new particle set  $\{X_k^{(i)}, (1/N)\}_{i=1}^N$ .

**3.1.2. Particle Filter Sampling Method.** In the process of particle filtering, the sampling method of particles will greatly affect the performance of the algorithm. Therefore, choosing a good sampling method is the key to the success of the overall algorithm. The Markov chain-Monte Carlo sampling (MCMC) method is to sample a specific sample  $X^{(i)}$  from the proposed distribution  $q(X)$  and obtain the state probability in the state space of  $X$  Markov chain of distribution  $p(X)$  so that high-dimensional distribution, invariant distribution, and Gibbs distribution are sampled through Markov chain samples, and the method of random simulation of sampled samples is Markov chain-Monte Carlo Sampling method. This sampling method is widely used in many aspects of random simulation evaluation [7]. The more important and outstanding sampling method is the Metropolis-Hastings (MH) algorithm, which was first proposed by Metropolis and later improved and improved by Hastings on its basis. It becomes the current MH algorithm. Its sampling method is

$$p_a(X, \tilde{X}) = \min \left\{ 1, \frac{p(\tilde{X})q(X|\tilde{X})}{p(X)q(\tilde{X}|X)} \right\}. \quad (19)$$

$X$  is the current system state,  $\tilde{X}$  is the sample value obtained by sampling,  $p(X)$  is the invariant distribution of the Markov chain,  $q(\tilde{X}|X)$  is the recommended distribution, and  $p_a(X, \tilde{X})$  is the acceptance probability of the required Markov chain to transfer to  $\tilde{X}$ .

The flow of the M-H algorithm is briefly described as follows:

- (1) Initialization: give the initial state  $X^{(0)}$  of the system.
- (2) Sample extraction: according to the previous system status  $X^{(i-1)}$ , candidate samples are sampled from the recommended distribution.
- (3) Acceptance probability: calculate the acceptance probability  $p_a(X, \tilde{X})$  according to  $p_a(X, \tilde{X}) = \min\{1, (p(\tilde{X})q(X|\tilde{X})/p(X)q(\tilde{X}|X))\}$ .
- (4) Calculating the state transition: generate a uniform random number  $u$  on the distribution of  $[0, 1]$ , compare the acceptance probability with  $u$ , if it is greater than  $u$ , and complete the state transition, namely,  $X^{(i)} = \tilde{X}$ ; otherwise, do not transfer, namely,  $X^{(i)} = X^{(i-1)}$ .

**3.1.3. Solution to the Problem of Particle Degradation.** In the particle filter algorithm described above, after multiple recursive operations, it is likely that only a few particles have large weights, while the rest of the particles have small weights. There are many particles with these small weights, and they have very little impact on the overall particle filter. In the calculation process, they take up a lot of time and greatly increase unnecessary computing time. In extreme cases, there may be only a few particles with very large weights left, while the weights of a large number of other particles are close to 0. Scholars call this phenomenon the degradation of particles [8].

The phenomenon of particle degradation is due to the narrow support set of the particle's proposed distribution  $q(X_k|X_{0:k-1}, Z_{1:k})$  and the particle's posterior distribution  $q(X_k|Z_{1:k})$ , and there is only a small amount of overlap. In the sampling calculation, only a small number of particles are concentrated in the support. In this way, after Bayesian estimation, the weight of these particles becomes very small. After many times of this process, the weight of the particles gets smaller and smaller and finally approaches 0. Since its calculations have not changed, it still takes a lot of time. In order to solve the problem of particle degradation, Liu et al. proposed the concept of effective particle number and defined a relation:

$$T_s \approx \sum_{i=1}^k w_k \rho_k. \quad (20)$$

Since the above formula is more complicated to calculate, the following simple formula is often used instead of the above formula:

$$\hat{N}_{\text{eff}} = \frac{1}{\sum_{i=1}^N (w_k^{(i)})^2}. \quad (21)$$

The degradation phenomenon of particles makes the weight of only a few particles larger, and the accuracy of using a few particles to express the probability distribution is low. Therefore, the degradation phenomenon of particles greatly compromises the reliability of particle filtering. How to avoid particle degradation or how to improve particle degradation had become the focus of research by many scholars. The more effective and efficient algorithm is the resampling method. This method is to add the resampling calculation after calculating the particle weight and resample the particles according to the particle weight. Particles with small weights are deleted to avoid taking up a lot of calculation time, and particles with large weights remain. At the same time, in order to ensure the number of particles, the particles with large weights are divided, and multiple sampling algorithms are used. The sampling variance is  $\text{Var}(N_i) = N w_k^{(i)} (1 - w_k^{(i)})$ . In order to improve the effectiveness of sampling, the sampling variance must be reduced. At this time, system sampling can be used. It is a sampling method with a very small sampling variance.

The process diagram of particle resampling is shown in Figure 6 as follows.

Figure 6 clearly illustrates the whole process from particle generation to weight change and resampling, that is, the overall implementation process from the posterior probability  $p(x_{k-1}|z_{1:k-1})$  at time  $k-1$  to the posterior probability density  $p(x_k|z_{1:k})$  at time  $k$ . The black solid dots represent the particles in the particle filter, and its size reflects the relationship between the particle weights. The particle weight represented by the large black dot is greater than the particle weight represented by the small black dot. The researcher initially assumed that there are ten particles, and the weights of these ten particles are all equal to  $(1/N)$ . The curve in the figure above represents the distribution state of all particles. The particles passing through the peak have a greater weight than those passing through the trough.

- (1) After ten particles with equal weights have passed the first curve, the weights of the ten particles have changed correspondingly due to the different positions of the crests and troughs of the passing curves. The weight of the particle passing through the wave crest is larger, and the weight of the particle passing through the trough is relatively smaller. The schematic diagram of the particle weight at  $k-1$  shown in the figure above is obtained.
- (2) After step 1, the weights of ten particles have changed. At this time, the resampling algorithm is used to keep the particles with larger weights. Particles with small weights are discarded and no longer participate in calculations, reducing the amount of subsequent calculations and leaving the number of effective particles.
- (3) After step 2, the particles with larger weights are retained, and they are repeatedly sampled to ensure that the number of particles does not decrease after sampling, and it is still ten particles. And the weights of ten particles are the same as the initial ones, and the prior probability density of particle  $k$  is obtained.
- (4) The number of resampled particles and the weight of each particle are exactly the same as the moment  $k-1$ . At this time, ten particles are observed to obtain the posterior probability density of the particles, and the entire process of particle resampling is completed.

Through the above calculation, the influence of small weight particles on the filtering algorithm is well eliminated. On the premise of ensuring the diversity of particles, the researchers remove particles with smaller weights, which greatly improves the utilization of calculations and the accuracy of filtering.

The aforementioned algorithm to avoid particle degradation can be briefly described as follows:

- (1) Weight calculation: for the sampled particle set  $\{\tilde{x}_k^{(i)}, w_k^{(i)}\}_{i=1}^N$ , set the initial weight  $c_1 = w_k^{(1)}$ , and use this to calculate the cumulative weight  $c_i = c_{i-1} + w_k^{(i)}$ .



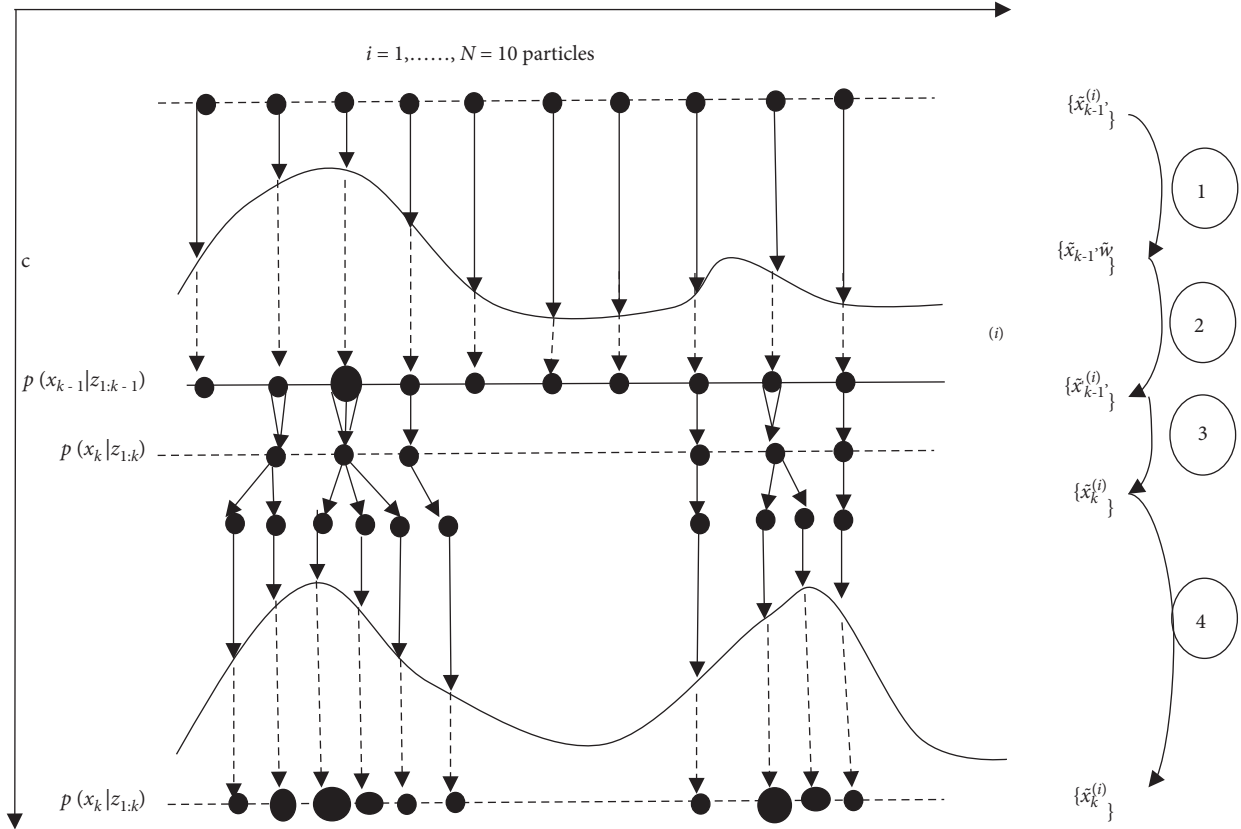


FIGURE 6: Schematic diagram of particle resampling.

- (2) Calculating the random number: calculate the uniformly distributed random number on  $[0, N^{-1}]$ , namely,  $u - U[0, N^{-1}]$ .
- (3) Resampling: for the range of  $j = 1: N$ ,  $u_1 = u + N^{-1}(j - 1)$ , if  $u_1 > c$ , then calculate the next weight; otherwise  $X_k^{(j)} = \tilde{X}_k^{(j)}$  and  $w_k^{(j)} = N^{-1}$ .

The flowchart of particle filtering is shown in Figure 7.

### 3.2. Vehicle Tracking Algorithm Based on Improved SPF

**3.2.1. Segmentation of Moving Vehicles.** Particle filter algorithm (SPF) is based on segmentation compensation [9]. In the process of tracking the target vehicle, this algorithm divides the target into multiple regions and selects the local features of the target from these regions. If the tracked target is partially occluded or morphologically changed during the tracking process, other areas of the target can be used instead of tracking, which can make the particle filter more accurate and applicable. After the target is located, the target to be tracked needs to be segmented, and the tracking frame is equally divided into the following nine regions as shown in Figure 8.

Similarly, segmentation methods such as  $2 \times 2$  or  $4 \times 4$  can also be selected, but  $2 \times 2$  segmentation has less effect on the segmentation of the target frame and loses the meaning of tracking frame segmentation. When the video resolution is not high, dense segmentation such as  $4 \times 4$  or even  $5 \times 5$

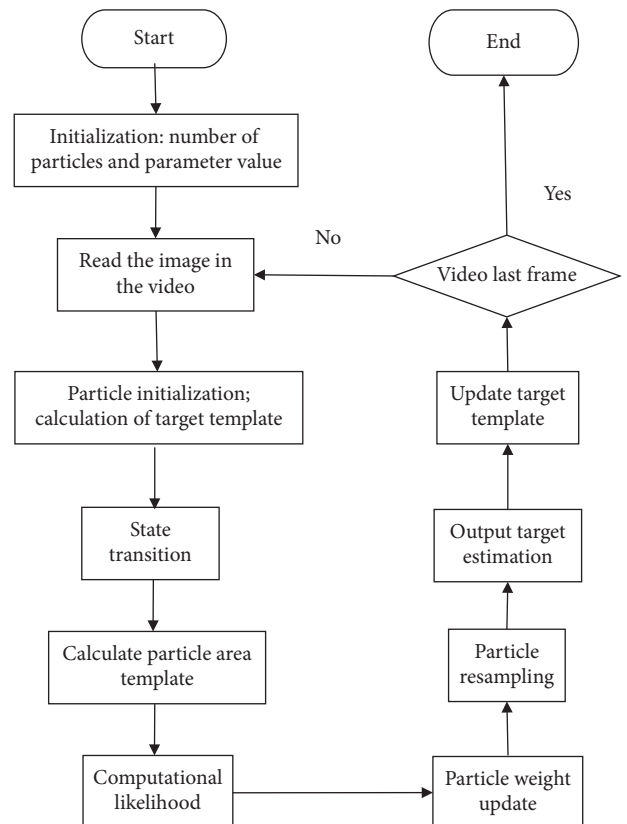


FIGURE 7: Particle filter target tracking flowchart.

|   |   |   |
|---|---|---|
| 1 | 2 | 3 |
| 4 | 5 | 6 |
| 7 | 8 | 9 |

FIGURE 8: Schematic diagram of target segmentation.

makes the resolution of the segmented area too small and fewer features are extracted from it, which is not conducive to the continuous tracking. The segmentation of the  $3 \times 3$  region selected in this paper is more practical and has no emphasis on the global average distribution. It can be applied to most target tracking situations, and the region with the best local features can be selected from these nine segments for vehicle track.

After the segmentation area is obtained, it is necessary to select a suitable segmentation area for tracking during the tracking process. For the 1-9 segmented regions defined in the above figure, the researcher defines the concept of segmentation adjacent. The adjacent bits of segment 1 are segment 2 and segment 4, and the adjacent bits of segment 5 are segment 2, segment 4, segment 6, and segment 8, and so on. After obtaining the relationship of adjacent bits, the selection rule for segmentation is defined as follows: the feature histogram of the segmented block and the feature histogram of adjacent bits are better if the Bhattacharyya coefficient is smaller. The Bhattacharyya coefficient is the similarity coefficient, which is defined as

$$\rho[\hat{p}(\hat{y}_0), \hat{q}] = \sum_{u=1}^m \sqrt{\hat{p}_u(\hat{y}_0) \hat{q}_u}. \quad (22)$$

The value range of its value is between 0 and 1, and the value of  $\rho$  reflects the similarity of the calculated two models. The larger the  $\rho$  value is, the higher the similarity of the two models. For the segmented region, the larger the  $\rho$  value of the adjacent bits, the higher the similarity between the two adjacent bits. In other words, the correlation is large and can be replaced by adjacent bits. Therefore, we take the segmentation area with a small  $\rho$  value, that is, the segmentation area with low similarity to the neighboring bits, and discard the segmentation area with a large  $\rho$  value.

After obtaining the segmentation area with small correlation, find the mean value of the Bhattacharyya coefficients of all the adjacent positions of the segmentation area; that is,

$$\rho_{\text{mean}} = \frac{\sum_i^k \rho_i}{k}. \quad (23)$$

For example, for segment 5, it is to find the Bhattacharyya coefficients of segment 2, segment 4, segment 6, and segment 8 and then add the sum and average.

After calculating the  $\rho$  values of all 9 segments, select the smaller segmentation area, and perform target tracking on the selected area.

*3.2.2. Motion Compensation for Segmented Areas.* Figure 9 is the particle distribution map formed by the segmented area at a certain time  $t$  during the vehicle tracking process. It can be clearly seen that the particles are distributed around the center of the segmented area.

The following figure shows the position A where the segmented area A moves at  $t + 1$  during the target tracking process. The particle distribution shown in the figure is still the particle distribution at time  $t$ .

Figure 10 shows the position where the segmented area A moves at  $t + 2$  during the target tracking process. The particle distribution shown in the figure is the particle distribution at  $t + 2$ .

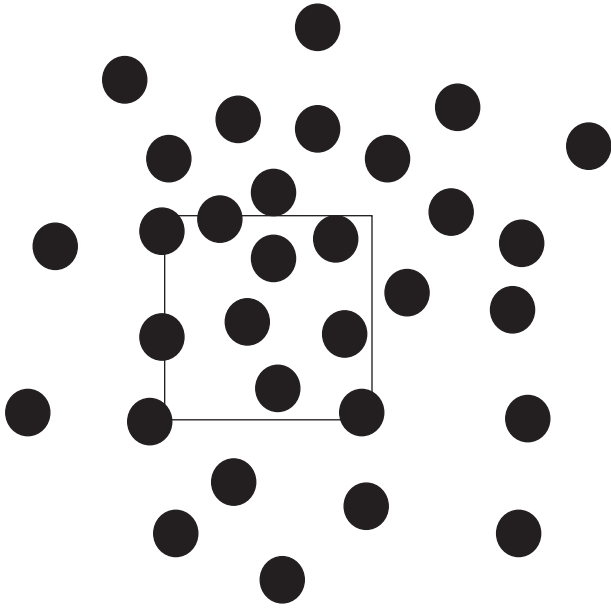
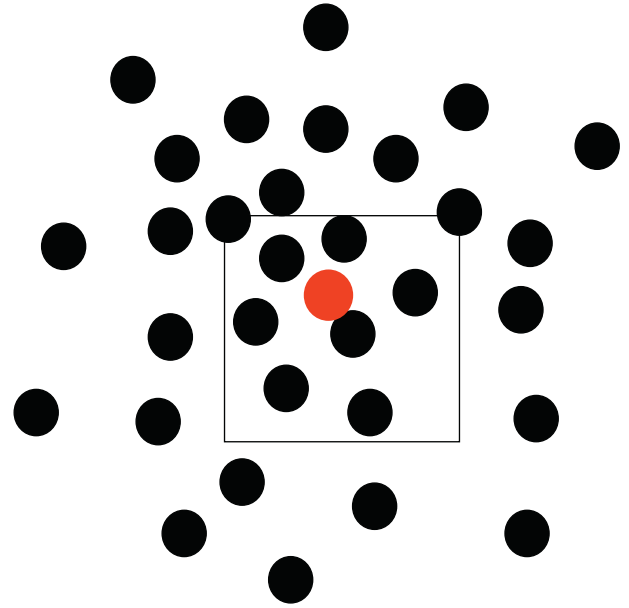
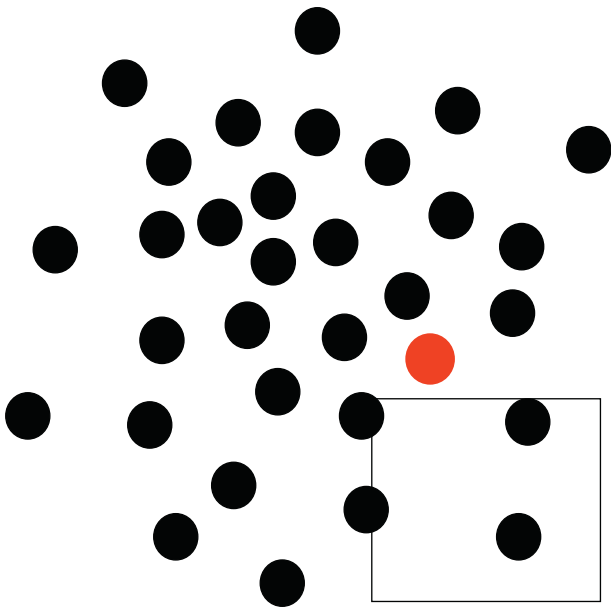
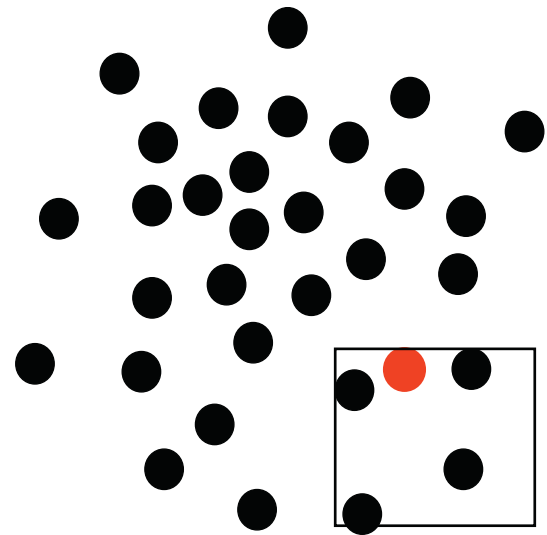
The situation in Figure 10 is exactly the same as that in Figure 11. Due to the faster movement of the segmented area, there are fewer particles at the location of the segmented area at  $t + 1$ , and the target is located at the position of the small red dot in Figure 11. At this time, if the researcher adds an offset to these particles to make the position of the particles shift to the direction of the target movement, the particles at  $t + 1$  can be shown in Figure 10, and the particle distribution at  $t + 1$  can be obtained. It is not the distribution centered on the position of the small red dot, but there is some deviation in the direction of the target movement, which is closer to the true position of the target in Figure 11. In this way, the segmented area tracked at  $t + 2$  after offset compensation is shown in (D).

Comparing the tracking frames in Figures 11 and 12, we can see that the target position is closer to the real one. The problem of tracking failure in Figure 10 does not occur.

The particle center is represented by the following formula:

$$P_p = P_p + \theta T_s. \quad (24)$$

$P_p$  represents the current position of the particle,  $T_s$  represents the movement displacement of the vehicle in the previous frame, and  $\theta$  represents the weighted value. The value of  $\theta$  is generally between 0.3 and 0.7 according to the movement form of the tracking target. In this paper, the median value is 0.5. In the application process, the particles scattered in the segmented area are added with an offset to improve the problem of large target positioning deviation caused by rapid target movement.

FIGURE 9: Particle distribution diagram at time  $t$ .FIGURE 11: Particle distribution diagram at time  $t + 2$ .FIGURE 10: Particle distribution diagram at time  $t + 1$ FIGURE 12: Particle distribution diagram at time  $t + 2$ .

### 3.2.3. Optimized Vehicle Tracking Based on SPF Algorithm.

Assume that the segmentation area with a smaller  $\rho$  value obtained in the previous section is segment 1, segment 5, and segment 9 and the center of the tracking frame is the black spot position, as shown in Figure 13.

Then, this paper uses the target tracking results of these three segmented regions to locate the overall target motion and obtains segment 1, segment 5, and segment 9 target tracking results as shown in Figure 14.

The final motion position of the target is the vector sum of segment 1, segment 5, and segment 9, which can be expressed by the following formula:

$$T_s \approx \sum_{i=1}^k w_k \rho_k. \quad (25)$$

$T_s$  represents the motion displacement of the overall target,  $\rho_k$  represents the motion displacement of the segmentation, and  $w_k$  is the weight of each segmentation. If there is no special target tracking situation,  $w_k$  can be set to the same value and the researcher only normalize the motion of all segmented regions. The final target tracking and positioning result is shown in Figure 15.

The advantage of such segmentation tracking is that, firstly, the segmented areas with greater correlation are eliminated, and they can be replaced with segmented areas

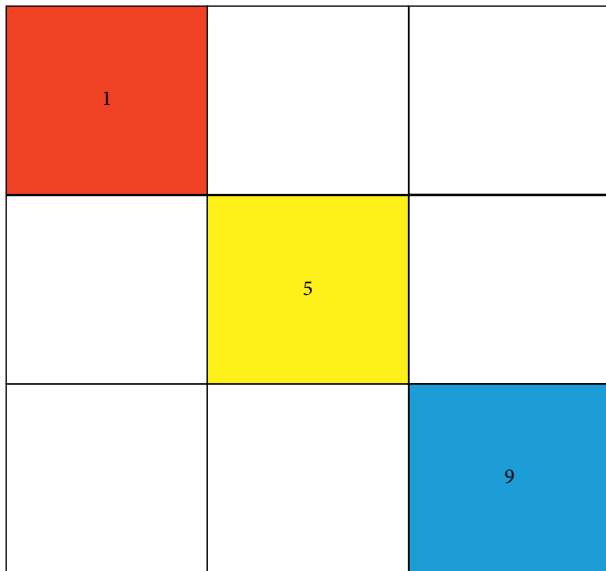


FIGURE 13: Schematic diagram of selecting partition area.

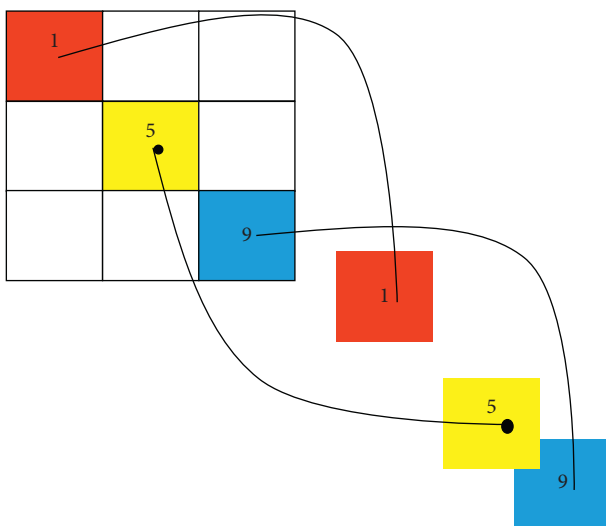


FIGURE 14: Schematic diagram of tracking results of segmented regions.

related to them, which reduces the amount of calculation accordingly. On the other hand, each small segmented area is equivalent to the target tracking of a small object, and the particle filter method is used to track it to ensure the accuracy of its tracking. In this way, the positive and negative errors of the target tracking result of the entire tracking frame obtained from the vector sum of the vehicle tracking results in a plurality of small segmented regions are offset, the tracking error is reduced, and the tracking accuracy is improved.

## 4. Test Analysis of Vehicle Tracking

*4.1. Comparison of Tracking Performance of Different Algorithms.* In order to verify the performance of the improved algorithm proposed in this chapter for vehicle

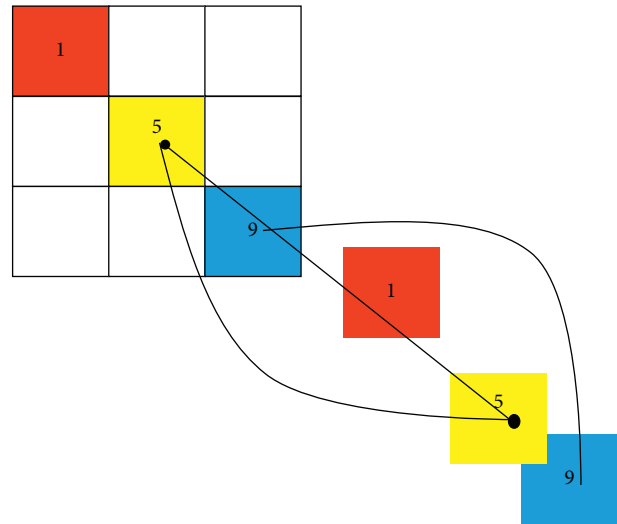


FIGURE 15: Schematic diagram of target tracking results.

tracking, this article uses the mainstream domestic Monte Carlo and convolutional neural network vehicle tracking algorithms to compare and first process the video 05 accident. The size of the video in Figure 16 is  $432 \times 240$ , and the number of frames is 125. This video is a video of a vehicle driving on a highway shot in a helicopter. In the video, a black car rolls over when it collides with another car while driving at a high speed. The tracking success rate, tracking accuracy, and algorithm time-consuming of the two algorithms are compared by tracking the target vehicle [10].

The tracking results of the three algorithms are shown in Figures 17–19, the red box represents the real tracking that passed the algorithm, and the blue box represents the false tracking that the algorithm failed to track.

The comparison results in the above figure can be divided into three sections to analyze:

- (1) Vehicle driving at high speed: Both Monte Carlo algorithm and SPF algorithm can accurately track vehicles, but SPF algorithm tracking is more accurate.
- (2) Vehicle collision: The tracked vehicle suddenly occludes and rotates, and at the same time, it is accompanied by the appearance of similar objects. The other two algorithms get wrong tracking results, while the SPF algorithm tracks accurately.
- (3) Vehicle out of control: Large-scale rotation and deformation of the vehicle. At this time, because similar vehicles leave the screen, the other two algorithms cannot track the target, the tracking frame turns blue, and the SPF algorithm still successfully tracks the target.

### 4.2. Analysis of Tracking Results

- (1) The vehicle is traveling at high speed: Both algorithms in the first segment of the video successfully track the target vehicle, but due to the high-speed movement of the target vehicle, the Monte Carlo



FIGURE 16: Original video frame.



FIGURE 17: The tracking results of the particle filter algorithm for segmentation compensation in this paper.



FIGURE 18: Monte Carlo tracking results.



FIGURE 19: Convolutional neural network vehicle tracking.

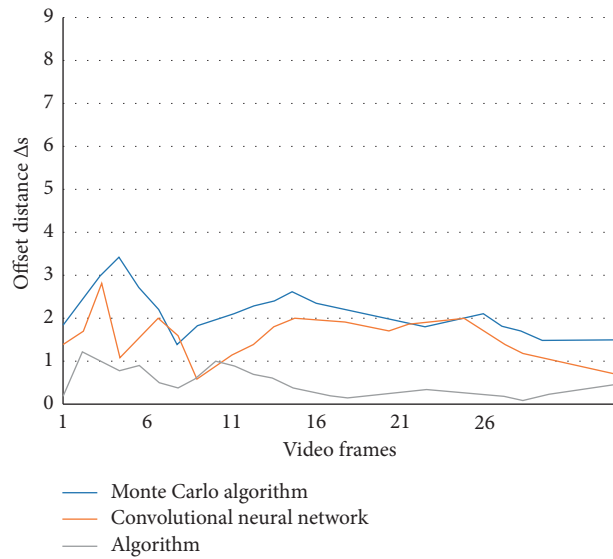


FIGURE 20: Comparison of vehicle offset distances with different algorithms.

TABLE 2: Average error comparison.

|  | Average error |
|--|---------------|
| Monte Carlo algorithm                  | 4.7204 pixels |
| SPF algorithm                          | 1.7354 pixels |
| Convolutional neural network algorithm | 3.6781 pixels |

algorithm is successful in tracking the target. The targets at  $t + 1$  are always scattered at the edge of the target area at  $t$ , leading to tracking deviation problems. In this case, there will be an error rate of tracking deviation of 4.7204.

This video has a total of 28 frames. The method in the previous section is used to compare the deviation between the target tracked by the three algorithms and the real target. The results are shown in Figure 20.

Table 2 also shows the advantages of the SPF algorithm proposed in this chapter in tracking high-speed moving vehicles, and the average error is reduced by 16.48%.

## 5. Conclusion

In regard to the road traffic environment, this article focuses on moving vehicle target tracking technology to complete the research. It is the future development trend to flexibly solve road traffic problems through intelligent transportation systems. Among them, the vehicle target tracking algorithm is one of the key technologies, which is also an important research direction of image processing and computer vision. Vehicle tracking technology is a subject that combines many advanced technological achievements such as image processing, pattern recognition, and automatic control.

The following innovations are worth noting:

- (1) Through the corresponding relationship between the vertex adjacency matrix and the regional behavior and vehicle behavior, this paper constructs the vertex adjacency matrix that reflects the characteristics of the undirected bipartite graph and simplifies the logic to complete the regional behavior analysis and vehicle behavior analysis.
- (2) By dividing the target area, this paper selects the local features of each part of the target and draws the conclusion. If the tracked target is partially occluded or changed in shape during the tracking process, other areas of the target can be used instead of tracking, which optimizes the accuracy and applicability of particle filtering.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no known conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This research was funded partially by the General Project of Science and Technology Program of Beijing Municipal Education Commission, Grant no. KM201911417011, Science and Technology Leading Fund of Smart City College, Project no. 11101611614, and Beijing Municipal Natural Science Foundation (Grant no. 4194078).

## References

- [1] X. Zhang and H. Xu, "Research on moving vehicle detection algorithm based on video processing," *Liquid Crystal Display*, vol. 12, no. 1, pp. 112–117, 2012.
- [2] M. Qi, Y. Pan, and Y. Zhang, "Preceding moving vehicle detection based on shadow of chassis," *Journal of Electronic Measurement and Instrument*, vol. 26, no. 1, pp. 54–59, 2012.
- [3] Z. Wang and H. Zeng, "New moving vehicle detection and tracking algorithm based on the combination of interframe difference and optical flow technology," *Computer Applications and Software*, vol. 15, no. 5, pp. 117–120, 2012.
- [4] R. Tan, Y. Lei, and B. Chen, "Vehicle detection at night," *Computer Engineering and Applications*, vol. 41, no. 13, pp. 227–228, 2005.
- [5] L. Cao, H. Zheng, L. Xi et al., "All-weather moving vehicle detection method based on multi-channel fusion of HOG features," *Journal of Wuhan University (Information Science Edition)*, vol. 40, no. 8, pp. 1048–1053, 2015.
- [6] X. Zhang and H. Xu, "Research on moving vehicle detection algorithm based on video processing," *Chinese Journal of Liquid Crystals & Displays*, vol. 27, no. 1, pp. 108–113, 2012.
- [7] X. Zhang, R. He, Z. Liu et al., "Moving vehicle detection based on spatial information Gaussian mixture model," *Journal of Jiangsu University: Natural Science Edition*, vol. 39, no. 4, pp. 19–22, 2011.
- [8] Q. Lu, Z. Wu, Y. Fan et al., "Moving vehicle detection method based on Gaussian mixture model," *Optoelectronics. Laser*, vol. 24, no. 4, pp. 751–757, 2013.
- [9] J. Ma, *Research on Moving Vehicle Detection and Tracking Technology Based on Video*, Beijing University of Technology, Beijing, China, 2020.
- [10] Y. Han, *Research on Moving Vehicle Detection and Tracking Algorithm Based on Video Images*, Beijing University of Technology, Beijing, China, 2011.

## Research Article

# Power Control for Full-Duplex Device-to-Device Underlaid Cellular Networks: A Stackelberg Game Approach

Zhen Yang,<sup>1</sup> Titi Liu ,<sup>2</sup> and Guobin Chen<sup>1</sup>

<sup>1</sup>Chongqing Key Laboratory of Spatial Data Mining and Big Data Integration for Ecology and Environment, Rongzhi College of Chongqing Technology and Business University, Chongqing 401320, China

<sup>2</sup>College of Medical Information Engineering, Shandong First Medical University and Shandong Academy of Medical Sciences, Taian 271016, China

Correspondence should be addressed to Titi Liu; [zwsun@sdfmu.edu.cn](mailto:zwsun@sdfmu.edu.cn)

Received 3 August 2020; Revised 30 September 2020; Accepted 10 October 2020; Published 11 November 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Zhen Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In spectrum sharing cognitive radio networks, unauthorized users (secondary users) are allowed to use the spectrum of authorized users (primary users) to improve spectrum utilization. Due to limited spectrum resources, how to formulate a reasonable spectrum allocation scheme is very important. As a mathematical analysis tool, game theory can solve the problem of resource allocation well. In recent years, it has been applied to the research of resource allocation in spectrum sharing networks by some literatures. In a cellular network consisting of multiple cellular users and full-duplex end-to-end communication users D2D (device-to-device), the self-interference caused by full-duplex communication and the interference caused by the D2D users to the cellular users will significantly reduce system throughput. In order to reduce the interference in the network, this paper introduces a power control algorithm based on Stackelberg game, which sets the cellular users and D2D users as the leaders and followers, respectively. The cellular users and the D2D users compete with each other to minimize the cost, and we propose new utility functions. We build an optimization problem under the outage probability constraint and power constraint and the transmission power of the users is obtained by using the Lagrangian dual decomposition method. The simulation results show that the proposed game algorithm improves network performance compared with other existing schemes.

## 1. Introduction

With the rapid development of 5G technology, more and more mobile devices are required to access, making spectrum resources even poorer. The traditional spectrum sharing scheme is a static spectrum allocation mode. It is to allocate some fixed frequency bands to users of some specific networks. These specific users are authorized users (primary users). Users in networks that have not been allocated spectrum resources are unauthorized users (secondary users). In the past, these unauthorized users (such as Wi-Fi users) could only use a small number of users.

Crowded public open unauthorized spectrum, while most authorized frequency bands are sometimes idle (such as TV frequency bands). In order to improve the utilization rate of radio spectrum and make good use of radio spectrum

resources reasonably, a dynamic sharing scheme must be proposed to allocate radio spectrum resources more efficiently. Therefore, cognitive radio (CR) technology is proposed in this environment. As a new concept, cognitive radio technology promotes the development of wireless communication industry, improves the utilization efficiency of spectrum resources, and provides a brand-new idea for realizing the sharing of spectrum resources.

The high transmission rate of 5G cellular network can effectively promote the realization of Internet of Things (IoT) and smart city. However, the big data brought by the Internet of Things and smart cities may overload 5G cellular networks due to limited spectrum resources, and mobile communication networks gradually cannot meet people's exponential demand for wireless mobile data traffic. D2D technology can relieve the pressure of the upcoming core

network traffic and is considered as a promising way to solve this problem. D2D communication allows direct communication between devices without using base stations for relay, and the transmission distance of D2D pairs is short. Spectrum efficiency can be significantly improved through spectrum multiplexing of cellular users. Therefore, D2D communication has the advantages of reducing the network burden of cellular users, reducing the power consumption of mobile terminals, improving the robustness of network infrastructure, improving spectral efficiency and system capacity, reducing user network delay and power consumption, expanding network coverage, etc. D2D technology can be used in machine-to-machine communication, car-to-car communication, content distribution, location-aware services in smart cities, social networks, proximity online games, electronic health care, and public safety services. However, in order to realize D2D communication, several challenging problems need to be solved, such as equipment discovery, spectrum resource allocation, interference management, power control, communication security, etc.

With the popularization of the fifth-generation mobile communication network and terminal equipment, the requirement for high-speed transmission of the network is becoming higher and higher. Traditional LET network technology cannot meet the requirements of high communication services, especially high-speed services. Thus, device-to-device (D2D) mode of direct communication between close-range terminals and terminals has been proposed in this context. This communication mode shares spectrum resources with cellular users in cellular networks to improve the capacity and spectrum utilization of network systems. It has become one of the most popular research technologies recently. At the same time, simultaneous full-duplex D2D communication technology can support two communication devices to transmit and receive at the same frequency band in the wireless cellular network, and further improve the spectrum utilization [1]. However, how to reduce power interference among users is a key problem, such as D2D user interference, D2D user interference to cellular users, and D2D user self-interference.

## 2. Related Work

In order to solve the resource management problem in D2D cellular network, some non-cooperative or cooperative game theory schemes are proposed in [2–5], which are used to control the power problem of DD users and cellular users in D2D cellular network, reduce the interference between users, and maximize the capacity of users in the network. In addition, the convergence of user transmit power is also analyzed in this paper. However, these works basically only consider half-duplex D2D communication mode. Therefore, the scheme of user power control and interference management in D2D full-duplex communication mode is proposed in [6–8]. Literature [6] proposes a rate maximization model to improve the total rate of full-duplex D2D communication network and minimize the interference of D2D users to cellular users. In [7, 8], a power optimization scheme

is proposed to maximize the total rate of cellular subscribers and D2D pairs in the presence of two directional interference. In addition, the paper [9, 10] proposes an efficient adaptive user power control scheme, which further reduces the interference between users and improves the system performance. In document [11], in order to reduce user interference, an interference cancellation scheme of cell network selection model is designed. However, most of the existing work neglects studying the power control of multiple cellular users and multiple D2D users in full-duplex D2D communication networks. In addition, the user's requirement for interruption probability has not been taken into account in the establishment of the model. Although reference [3] considers the interruption probability constraints of the network, it only considers the half-duplex communication mode, and its utility function does not consider the network power saving in order to maximize the network throughput. Zulhasnine et al. proposed a greedy algorithm based on the maximum throughput of the system and proposed an optimization strategy from two aspects of D2D communication users multiplexing the uplink and downlink resources of cellular users. The allocation problem of D2D to users' cell radio resources was summarized as a mixed integer nonlinear programming problem [12]. Min et al. proposed a D2D mode selection scheme for users in downlink, which reduced the interference between communication users [13]. Yu et al. proposed a joint mode selection and resource allocation algorithm, which achieves the effects of maximum capacity and minimum communication interference of the communication system [14]. Xiao et al. studied the algorithm of D2D on the transmission power of users and cellular users and proposed a scheme to adjust the transmission power of users under different communication modes to control the interference between communication users [15]. Hassan et al. proposed a two-stage resource allocation algorithm to minimize system interference while maintaining the total rate of the target system. Firstly, a weighted two-part matching algorithm is used to minimize interference and obtain a feasible initial solution, and then local search technology is used to improve the solution [16]. Nguyen et al. and others studied from the geographical location of the cell where the users are located, established communication restricted areas according to the interference between users, and implemented different communication mode strategies in different areas, thus improving the overall communication quality of the system [17]. Ningombam and Shin proposed a radio resource allocation and power control scheme to reduce D2D interference to multiplexed uplink cellular resources [18]. A radio resource allocation scheme is proposed by using a cell sectorization method to reduce interference in D2D communication of LTE-A uplink cellular network. Jaheon et al. proposed a scheme combining resource allocation and power control. The scheme firstly coordinates the interference introduced by D2D to a certain extent by allocating reasonable multiplexing resources for D2D users and appropriately adjusts the transmission power of D2D users to ensure controllable cellular link interference [19]. Fodor et al. proposed a solution to share spectrum resources



between cellular mobile devices and D2D, thus improving the spectrum efficiency and energy efficiency of traditional cellular networks [20]. Doppler et al. explained the session and management mechanism of D2D communication and how to limit the interference of D2D communication to cellular networks [21]. Kim et al. proposed a D2D link radio resource allocation strategy for interference between D2D users and cellular users [22]. Li et al. proposed a radio resource allocation scheme to minimize the interference between communication users [23, 24], aiming at the interference problem generated in the communication system where D2D communication users coexist with cellular users and D2D technology is combined with Multi-User Multiple-Input Multiple-Output (MU-MIMO) technology.

Therefore, this paper proposes a power control algorithm for multiple cellular users and D2D users in full-duplex D2D communication with channel interruption constraints based on Stackelberg game. Considering the interference and self-interference in two directions among D2D users, we set cellular users and D2D users as leaders and followers, respectively. Cellular users and D2D users compete with each other to minimize their cost effectiveness (network power consumption), and propose a new utility function. Under the condition of channel interrupt constraint and power constraint, the optimization problem is constructed, and the transmission power of users is obtained by using Lagrange dual decomposition method. The simulation results show that the proposed game algorithm improves network throughput and reduces bit error rate. The second part introduces related work; the third part introduces the system model; the fourth part introduces the Stackelberg game theory method; the fifth part introduces the Stackelberg game solution; the sixth part introduces simulated experimentally.

### 3. System Model

The system model of cellular network is shown in Figure 1. A single base station is located in the cell center of the network. It has  $K$  cellular users. Cellular users have the priority to use the base station to allocate the channel.  $M$  D2D user's pair and D2D users pair are allowed to share the communication channel of cellular users, and D2D users adopt full-duplex communication mode. Suppose each D2D user has two antennas, one for transmitting signals and the other for receiving signals. Define the path loss and multipath fading channel coefficients of the cellular subscriber  $K$  and its receiver between the base stations as  $L_{mk}$  and  $h_{kk}$ ; for the path loss and multipath fading between the  $m$ -th D2D user pair and the  $k$ -th cellular subscriber and the base station, respectively, the channel coefficients are  $L_{mk}$  and  $h_{mk}$ , respectively; the path loss and multipath fading channel coefficients between the  $m$ -th D2D pair are  $L_m$  and  $h_{mm}$ ,

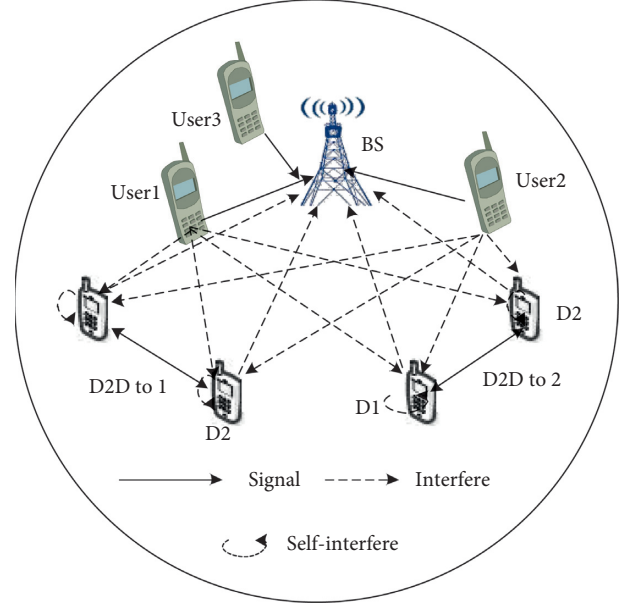


FIGURE 1: Cellular network model.

respectively, and for the path loss and multipath fading channel between the  $m$ -th D2D pair and the  $i$ -th user, the coefficients are  $L_{im}$  and  $h_{im}$ , respectively. The fading coefficient obeys the  $CN(0, 1)$  independent Gaussian distribution. Cellular user  $k$  has a transmit power of  $P_k$ . On the  $k$ -th cellular user channel, the powers of the first user (D1) and the second user (D2) of the  $m$ -th D2D pair are  $p_m^k$  and  $\alpha_m p_m^k$ , respectively, where  $\alpha$  is defined as the coefficient of the transmit power between D1 and D2. We assume that each cellular user transmits simultaneously on a single channel and that each cellular user's channel can simultaneously transmit all D2D pairs to the user.

The signal-to-interference and noise ratio (SINR) received by the  $k$ -th cellular subscriber at the base station is defined as

$$\gamma_k = \frac{p_k L_k |h_{kk}|^2}{\sum_{m=1}^M (1 + \alpha_m) p_m^k L_{mk} |h_{mk}|^2 + N_0}, \quad (1)$$

where  $L$  represents the path loss factor and  $L = d^{-3}$ , where  $d$  represents the distance between the user's transmitter and receiver, and  $N_0$  represents the noise power.

Since the current self-interference suppression technology cannot completely eliminate self-interference [9], the power of residual self-interference is defined as  $p_{SI} = \eta p$ , where  $p_{SI}$  is the power of residual self-interference,  $\eta$  ( $0 \leq \eta \leq 1$ ) is the cancellation coefficient of self-interference, and  $p$  represents the transmitting power of the transmitting end. On the  $k$ -th cellular user channel, the SINR received by D2 is defined as

$$\gamma_m^{k,D1 \rightarrow 2} = \frac{p_m^k L_m |h_{mm}|^2}{\sum_{i \neq m}^M (1 + \alpha_i) p_i^k L_{im} |h_{im}|^2 + \eta_{m,2} \alpha_m p_m^k L_{m,SI} |h_{m,SI}|^2 + p_k L_{km} |h_{km}|^2 + N_0}, \quad (2)$$

where  $L_{m,SI}$  and  $h_{m,SI}$  represent the self-interference path loss and fading channel coefficients of the  $m$ -th D2D pair to the user, respectively. Similarly, we can get the SINR of D2 to D1 on the  $k$ -th cellular user channel as  $\gamma_m^{k,D_2 \rightarrow 1}$ .

When the cellular user  $K$ 's SINR value is lower than the threshold  $\bar{\gamma}_k$ , its communication will be interrupted, and the cellular user's outage probability  $P_{k,out}^C(p_k, \mathbf{P}_{-k})$  is expressed as [13]

$$P_{k,out}^C(p_k, \mathbf{P}_{-k}) = 1 - \exp\left(-\frac{\bar{\gamma}_k N_0}{P_k L_k}\right) \prod_{m=1}^M \left(1 + \frac{\bar{\gamma}_k (1 + \alpha_m^k) P_m^k L_{mk}}{P_k L_k}\right)^{-1}, \quad (3)$$

where  $\mathbf{P}_{-k}$  represents the power vector of the user other than the cellular user on the channel of the  $k$ -th cellular user. Since the cellular user  $k$  has the constraint  $\xi_k$  of the channel outage probability,  $P_{k,out}^C \leq \xi_k$ , it can be obtained by formula (3):

$$1 - \xi_k \leq \exp\left(-\frac{\bar{\gamma}_k N_0}{P_k L_k}\right) \prod_{m=1}^M \left(1 + \frac{\bar{\gamma}_k (1 + \alpha_m^k) P_m^k L_{mk}}{P_k L_k}\right)^{-1}. \quad (4)$$

Similarly, on the channel of the cellular user  $k$ , when the SINR value  $\gamma_m^k$  of the D2D to the user is less than the target threshold  $\bar{\gamma}_m^k$  of the SINR, the probability of the interruption of the  $m$ -th D2D pair (from D1 to D2) can be defined as

$$P_{m,out}^{D_1 \rightarrow 2}(p_m^k, \mathbf{P}_{-m}^k) = 1 - \exp\left(-\frac{\bar{\gamma}_m^k (N_0 + \eta_{m,2} \alpha_m P_m^k L_{m,SI} |h_{m,SI}|^2)}{P_m^k L_m}\right) \cdot \prod_{i \neq m}^M \left(1 + \frac{\bar{\gamma}_m^k (1 + \alpha_i) P_i^k L_{im}}{P_m^k L_m}\right)^{-1}, \quad (5)$$

where  $\mathbf{P}_{-k}$  represents the transmit power vector of all other users except the  $k$ -th D2D on the first channel. If the probability of break  $\gamma_{m,out}^{k,D_2 \rightarrow 1}$  is less than the interrupt threshold  $\xi_m$ ,  $\gamma_{m,out}^{k,D_2 \rightarrow 1} \leq \xi_m$ , the constraint of the probability of interruption can be obtained:

$$1 - \xi_m \leq 1 - \exp\left(-\frac{\bar{\gamma}_m^k N_0}{P_m^k L_m} - x_m^{D_1 \rightarrow 2}\right) \prod_{i \neq m}^M \left(1 + \frac{\bar{\gamma}_m^k (1 + \alpha_i) P_i^k L_{im}}{P_m^k L_m}\right)^{-1}, \quad (6)$$

where  $x_m^{D_1 \rightarrow 2} = ((\bar{\gamma}_m^k \eta_{m,2} \alpha_m L_{m,SI} |h_{m,SI}|^2) / L_m)$ .

Similarly, the probability of the first D2D pair (from D2 users to D1 users) can be defined as

$$1 - \xi_m \leq 1 - \exp\left(-\frac{\bar{\gamma}_m^k N_0}{\alpha_m P_m^k L_m} - x_m^{D_2 \rightarrow 1}\right) \prod_{i \neq m}^M \left(1 + \frac{\bar{\gamma}_m^k (1 + \alpha_i) P_i^k L_{im}}{\alpha_m P_m^k L_m}\right)^{-1}, \quad (7)$$

where  $x_m^{D_2 \rightarrow 1} = ((\bar{\gamma}_m^k \eta_{m,1} L_{m,SI} |h_{m,SI}|^2) / (\alpha_m L_m))$ .

## 4. Stackelberg Game Theory

**4.1. Traditional Noncooperative Game Model.** In cognitive radio network system, it is considered that secondary users are access users who will not cooperate with each other; then there must be conflicts between secondary users, which plays an extremely important role in analyzing complex and competitive cognitive radio networks with noncooperative game theory. Let  $\Theta = \{\kappa, \{P_k\}\{U_k(p)\}\}$  be NPG model and  $\kappa = [1, 2, 3, \dots, K]$  be the set of perceived users.  $P_k$  is the power set of secondary user  $k$ ; its revenue function is  $U_k(*)$ . In order to solve the problem that the game model cannot obtain the global optimal transmission power solution, a linear cost function is introduced into the utility function of the secondary user, and the benefit function expression of NPGP model [25] is defined as follows:

$$\text{NPGP: } \max_{p_k \in P_k} U_k(p_k, P_{-k}) = \frac{\text{LR}}{M P_k} \left(1 - e^{(-\gamma_k/2)}\right)^M - c_1 p_k, \quad k \in \kappa. \quad (8)$$

In the formula,  $M$  is the total number of bits per frame;  $L$  is that number of valid information bit ( $L$  & lt;  $M$ );  $R$  is transmission rate;  $c_1$  is a penalty factor;  $P$  is secondary user transmit power vector:  $p = (p_1, \dots, p_k)$ ,  $P_{-k}$  is the power of other users besides perceiving user  $k$ .

Utility function  $U_k$  represents the final income of users after the game in bits/joules. However, the above penalty factor is a constant, and due to different interferences caused by users, it is unfair to set it in this way. Through the introduction of the above-mentioned related game schemes, the improved penalty function [26] is expressed as follows:

$$\text{NPGP - MSFLA: } U_k(p_k, P_{-k}) = \frac{\text{LR}}{M P_k} \left(1 - e^{(-\gamma_k/2)}\right)^M - c_1 e^{p_k} - c_3 (\gamma_k - \gamma_k^{\min}). \quad (9)$$

In the formula,  $c_2$  and  $c_3$  are constant cost factors.

Let  $f_1(\gamma_k)$  be an effective function when data transmission is successful. Generally, it is modulated by frequency shift keying. The expression is defined as follows:

$$f_1(\gamma_k) = \left(1 - e^{(-\gamma_k/2)}\right)^M. \quad (10)$$

In the game model of [27], a new benefit function is designed, in which the penalty function is a nonlinear function set according to the interference of users and the signal-interference-noise ratio, and the utility function is defined.

$$\text{NPGP - ESIA: } U_k(p_k, P_{-k}) = \frac{\text{LR}}{M P_k} \left(\frac{1 - e^{-\gamma_k}}{1 + e^{\gamma_k - \gamma_k^{\min}}}\right) - \alpha e^{\beta((\gamma_k / \gamma_k^{\min}) - 1)} \frac{P_k}{p^{\text{th}}}. \quad (11)$$

In the formula,  $\alpha$  is that trim factor of the penalty function and is constant;  $\beta$  is that trim factor of the penalty function and is constant;  $p^{\text{th}}$  is threshold of average power.

**4.2. Stackelberg Game Model.** Stackelberg's game model is actually an orderly game model; that is, some actors can act first and are called leaders. However, the users who act later make corresponding action strategies according to the users who act earlier, which are called followers. This is the Stackelberg game. At the top of the game, leaders will choose the best action to maximize their benefits when they know the relevant information of followers. At the bottom of the game, followers make the best strategy through noncooperative game under the condition of understanding the relevant behaviors of leaders. The Nash equilibrium solution of Stackelberg game is called Stackelberg equilibrium (SE). The game leader will formulate his own optimal response function considering the follower's response information, which is defined as follows:

$$S_F^* \leq \arg U(S_L). \quad (12)$$

Leaders get their best action strategy  $S_F^*$  when maximizing their own benefit function. At this time, followers choose the best action strategy  $S_F^*$  according to the leader's response  $S_L^*$  to maximize their own interests.

$$S_L^* \leq \arg U(S_L). \quad (13)$$

After many games, Stackelberg equilibrium solutions ( $S_F^{\text{SE}}, S_L^{\text{SE}}$ ) can be obtained. Both parties change their strategies according to the information of other participants to further improve their income. Stackelberg equilibrium solution: the behavioral strategy set ( $S_F^{\text{SE}}, S_L^{\text{SE}}$ ) is the strategic equilibrium solution of Stackelberg's game. For any leader  $\forall i \in K$ ,

$$\begin{aligned} U_i(S_1, \dots, S_i^{\text{SE}}, \dots, S_K, \dots, S_L^{\text{SE}}) \\ \geq U_i(S_1, \dots, S_i, \dots, S_K, \dots, S_L^{\text{SE}}), \quad \forall i. \end{aligned} \quad (14)$$

For any one of the followers  $\forall j \in N$ ,

$$\begin{aligned} U_i(S_1, \dots, S_j^{\text{SE}}, \dots, S_K, \dots, S_F^{\text{SE}}) \\ \geq U_i(S_1, \dots, S_j, \dots, S_K, \dots, S_F^{\text{SE}}), \quad \forall j. \end{aligned} \quad (15)$$

The power control model can adopt Stackelberg game algorithm to control the transmission power of cognitive users according to the different order of primary and secondary users' strategies. In addition, since most literatures do not consider the outage probability constraints of primary users and secondary users into the system model, taking the constraints into account will improve the practicability of the algorithm, and Lagrange optimization method and KTT scheme are used to obtain the optimal transmission power of each user on the corresponding channel.

The power control strategy between cellular users and D2D to user can use the Stackelberg game model to solve power control problems. Specifically, cellular users are leaders, first taking action, setting the market price to sell the spectrum resources they use; D2D pairs are followers, observing the leader's actions and taking appropriate actions, by paying the main user's corresponding fees to share the licensed band of cellular users. Both cellular users and D2D pairs are selfish, and they are designed to minimize their own costs under outage constraints, thereby building a corresponding benefit function.

We first analyze the leader subgame of cellular user power control issues. Cellular users aim to minimize their cost by providing interference prices to D2D pairs. The cost utility of the leader (cellular users) can be defined as their own transmit power minus the gain from the D2D pair by sharing resources with all D2D pairs. The utility cost function for cellular users is expressed as

$$Q_k(p_k, \mathbf{p}_{-k}, c_k) = p_k - c_k \sum_{m=1}^M \left( (1 + \alpha_m) p_m^k L_{mk} |h_{mk}|^2 \right), \quad (16)$$

where  $c_k$  represents the interference price of the cellular user  $k$ .

The optimization issues for cellular users are as follows:

$$\begin{aligned} \min Q_k(p_k, \mathbf{p}_{-k}, c_k) \\ \text{s.t.} \quad (4). \end{aligned} \quad (17)$$

For a follower's subgame, the cost function for each D2D pair is the total power allocated on the channel of all cellular users plus the cost paid for using the cellular user's channel. Therefore, the cost function of the first D2D to the user can be expressed as

$$\begin{aligned} U_m(\mathbf{p}_m, \mathbf{p}_{-m}, \mathbf{c}) = \sum_{k=1}^K \left( (1 + \alpha_m) p_m^k \right) \\ + \sum_{k=1}^K \left( c_k (1 + \alpha_m) p_m^k L_{mk} |h_{mk}|^2 \right), \end{aligned} \quad (18)$$

where  $\mathbf{p}_m$  represents the power vector of the  $m$ -th D2D pair to users on all  $k$ -cellular user channels,  $\mathbf{p}_{-m} = (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_{m-1}, \dots, \mathbf{p}_{m+1}, \dots, \mathbf{p}_M)$  represents the transmit power matrix of all other D2D pairs, and  $\mathbf{c} = (c_1, c_2, \dots, c_k)$  represents the price vector that the  $m$ -th D2D user pays to all cellular users.

Therefore, the first D2D pair is intended to solve the problem of multi-channel power control. The specific problems are

$$\begin{aligned} \min U_m(\mathbf{p}_m, \mathbf{p}_{-m}, \mathbf{c}) \\ \text{s.t.} \quad (6) (7). \end{aligned} \quad (19)$$

Since constraints (4), (6), and (7) are nonconvex, the games (17) and (19) under constraints are nonconvex. Therefore, we can optimize problems (17) and (19) by using

the logarithm of the variable:  $\tilde{p}_k = \log(p_k)$ ,  $\tilde{p}_m^k = \log(p_m^k)$ . Therefore, the optimization problems in (17) and (19) are translated into

$$\min Q_k(e^{\tilde{p}_k}, e^{\tilde{p}_{-k}}, c_k) = e^{\tilde{p}_k} - c_k \sum_{m=1}^M \left( (1 + \alpha_m) e^{\tilde{p}_m} L_{mk} |h_{mk}|^2 \right), \quad \forall k \quad (20)$$

$$\text{s.t.} \quad \frac{\bar{Y}_k N_0}{e^{\tilde{p}_k} L_k} + \sum_{m=1}^M \left( \log \left( \frac{\bar{Y}_k (1 + \alpha_m) e^{\tilde{p}_m} L_{mk}}{e^{\tilde{p}_k} L_k} \right) \right) \leq \log \left( \frac{1}{1 - \xi_k} \right), \quad \forall k,$$

$$\min U_m(e^{\tilde{p}_m}, e^{\tilde{p}_{-m}}, \mathbf{c}) = \sum_{k=1}^K \left( (1 + \alpha_m) e^{\tilde{p}_m} \right) + \sum_{k=1}^K \left( c_k (1 + \alpha_m) e^{\tilde{p}_m} L_{mk} |h_{mk}|^2 \right), \quad \forall m$$

$$\frac{\bar{Y}_m^k N_0}{e^{\tilde{p}_m} L_m} + x_m^{D_1-2} + \sum_{i \neq m}^M \left( \log \left( 1 + \frac{\bar{Y}_m^k (1 + \alpha_i) e^{\tilde{p}_i} L_{im}}{e^{\tilde{p}_m} L_m} \right) \right) \leq \log \left( \frac{1}{1 - \xi_m} \right), \quad \forall k \quad (21)$$

$$\text{s.t.} \quad \frac{\bar{Y}_m^k N_0}{\alpha_m e^{\tilde{p}_m} L_m} + x_m^{D_2-1} + \sum_{i \neq m}^M \left( \log \left( 1 + \frac{\bar{Y}_m^k (1 + \alpha_i) e^{\tilde{p}_i} L_{im}}{\alpha_m e^{\tilde{p}_m} L_m} \right) \right) \leq \log \left( \frac{1}{1 - \xi_m} \right), \quad \forall k.$$

In the system herein, the cellular user's channel is fixed. For example, each cellular user occupies one subchannel, and then the D2D user selects whether to share the cellular user's channel for communication according to the cellular user channel condition. Therefore, this paper introduces the Stackelberg game theory to control the user power and manage the D2D user access cellular user channel; that is, the cellular user performs its own transmit power according to its own utility function (equation (16)), while the D2D user selectively accesses the cellular user's channel and controls its own transmit power according to its utility function (equation (18)).

The solution to the Stackelberg game is to analyze the existence and uniqueness of the Nash equilibrium. The tree constraint functions of (20) and (21) have  $\tilde{P}_{m,\text{out}}^c(\tilde{p}_k, \tilde{p}_{-k})$ ,  $\tilde{P}_{m,\text{out}}^{D_1-2}(\tilde{p}_m^k, \tilde{p}_{-m}^k)$ , and  $\tilde{P}_{m,\text{out}}^{D_2-1}(\tilde{p}_m^k, \tilde{p}_{-m}^k)$ . The performance problem of Nash Equilibrium can be expressed as follows.

**Theorem 1.** *The subgame of D2D to the user (follower) is the existence of Nash equilibrium and is unique.*

*Proof.* For each D2D pair,  $\forall m \in M$ , the interrupt constraint functions  $\tilde{P}_{m,\text{out}}^{D_1-2}(\tilde{p}_m^k, \tilde{p}_{-m}^k)$  and  $\tilde{P}_{m,\text{out}}^{D_2-1}(\tilde{p}_m^k, \tilde{p}_{-m}^k)$  are strictly convex functions and are monotonically decreasing in  $\tilde{p}_m^k$ . The power vectors  $\mathbf{P}^k = \{p_m^k; \forall m \in M\}$  in the Hessian matrices  $\tilde{P}_{m,\text{out}}^{D_1-2}$  and  $\tilde{P}_{m,\text{out}}^{D_2-1}$  are both non-negative. Therefore, problem (21) is obviously an optimization problem of strictly convex constraints, NE always exists, and there is a unique optimal solution.

**Theorem 2.** *The subgame of a cellular user (leader) is the existence of Nash equilibrium and is unique.*

*Proof.* For each cellular user, the  $\forall k \in K$ , the  $k$ -th cellular user's interrupt constraint function is  $\tilde{P}_{m,\text{out}}^c(\tilde{p}_k, \tilde{p}_{-k})$ ; the interrupt constraint function is a strict convex function and is about the  $\tilde{p}_m^k$  monotonically decreasing function. Therefore, problem (20) is an optimization problem of strict convex constraints, and there is a unique optimal solution.

The Nash equilibrium of the Stackelberg game consists of the Nash equilibrium of the follower subgame and the Nash equilibrium of the leader subgame and analyzes the existence and uniqueness of the Nash equilibrium of the Stackelberg game. The Nash equilibrium of the leader's subgame exists and is unique. For a given power allocation scheme, the Nash equilibrium of the follower subgame exists and is unique.

## 5. Stackelberg Game Solution

Since the optimal power control problems of (20) and (21) have constraints, if the complexity of solving the subproblem is higher with the distributed algorithm, we use the Lagrangian dual decomposition method to solve problem (20). And (21) optimizes the problem with the most transmit power.

**5.1. D2D Solves the User's Optimal Transmit Power.** D2D minimizes its cost utility by observing the interference price vector  $\mathbf{c}$  set by all cellular users and controlling its own transmit power. Thus, the Lagrangian function of equation (22) can be expressed as

$$\begin{aligned}
L_m(\tilde{p}_m, \tilde{p}_{-m}, \mathbf{c}, \lambda_m, \omega_m) &= \sum_{k=1}^K \left( (1 + \alpha_m) e^{\tilde{p}_m^k} \right) + \sum_{k=1}^K \left( c_k (1 + \alpha_m) e^{\tilde{p}_m^k} L_{mk} |h_{mk}|^2 \right) \\
&+ \sum_{k=1}^K \lambda_m \cdot \left( \frac{\tilde{\gamma}_m^k N_0}{e^{\tilde{p}_m^k} L_m} + x_m^{D_1-2} + \sum_{i \neq m}^M \left( \log \left( 1 + \frac{\tilde{\gamma}_m^k (1 + \alpha_i) e^{\tilde{p}_i^k} L_{im}}{e^{\tilde{p}_m^k} L_m} \right) \right) - \log \left( \frac{1}{1 - \xi_m^k} \right) \right) \\
&+ \sum_{k=1}^K \omega_m \left( \frac{\tilde{\gamma}_m^k N_0}{\alpha_m e^{\tilde{p}_m^k} L_m} + x_m^{D_2-1} + \sum_{i \neq m}^M \left( \log \left( 1 + \frac{\tilde{\gamma}_m^k (1 + \alpha_i) e^{\tilde{p}_i^k} L_{im}}{\alpha_m e^{\tilde{p}_m^k} L_m} \right) \right) - \log \left( \frac{1}{1 - \xi_m^k} \right) \right),
\end{aligned} \tag{22}$$

where  $\lambda_m$  and  $\omega_m$  are the dual variables of the constraints. The optimal solution for the Lagrangian dual function is  $\min L_m(\tilde{P}_m, \tilde{P}_{-m}, \mathbf{c}, \lambda_m, \omega_m)$ . According to the convex optimization theory,  $L_m(\tilde{P}_m, \tilde{P}_{-m}, \mathbf{c}, \lambda_m, \omega_m)$  is a concave

function with respect to  $\lambda_m$  and  $\omega_m$ ; therefore, the optimal dual variable that can be obtained by the subgradient method is as follows:

$$\begin{aligned}
\lambda_m^* &= \lambda_m^* + \varepsilon \left( \frac{\tilde{\gamma}_m^k N_0}{e^{\tilde{p}_m^k} L_m} + x_m^{D_1-2} + \sum_{i \neq m}^M \left( \log \left( 1 + \frac{\tilde{\gamma}_m^k (1 + \alpha_i) e^{\tilde{p}_i^k} L_{im}}{e^{\tilde{p}_m^k} L_m} \right) \right) - \log \left( \frac{1}{1 - \xi_m^k} \right) \right)^+, \\
\omega_m^* &= \omega_m^* + \phi \left( \frac{\tilde{\gamma}_m^k N_0}{\alpha_m e^{\tilde{p}_m^k} L_m} + x_m^{D_2-1} + \sum_{i \neq m}^M \left( \log \left( 1 + \frac{\tilde{\gamma}_m^k (1 + \alpha_i) e^{\tilde{p}_i^k} L_{im}}{\alpha_m e^{\tilde{p}_m^k} L_m} \right) \right) - \log \left( \frac{1}{1 - \xi_m^k} \right) \right)^+,
\end{aligned} \tag{23}$$

where  $\varepsilon$  and  $\phi$  represent small enough steps and are positive.

Since the  $L_m(\tilde{P}_m, \tilde{P}_{-m}, \mathbf{c}, \lambda_m, \omega_m)$  function is a concave function on  $\lambda_m$  and  $\omega_m$ , (15) iteratively converges to the optimal  $\lambda_m$  and  $\omega_m$ . Substituting the obtained optimal  $\lambda_m$  and  $\omega_m$  into (23), we derive it to obtain

$$\frac{\partial L_m(\tilde{P}_m, \tilde{P}_{-m}, \mathbf{c}, \lambda_m, \omega_m)}{\partial \tilde{p}_m^k} \Big|_{\lambda_m^*, \omega_m^*} = 0. \tag{24}$$

By solving (24), we can derive the optimal solution  $\tilde{p}_m^{k*}$  in (22). Then, by the exponential change of the  $p_m^{k*} = \exp(\tilde{p}_m^{k*})$ , the optimal transmit power  $p_m^{k*}$  of D1 in the  $m$ -th D2D pair and the optimal transmit power  $\alpha_m p_m^{k*}$  of D2 can be obtained. Therefore, we can get the optimal transmit power of each D2D to the user.

## 5.2. Cellular User (Leader) Optimal Transmit Power Solution.

In order to minimize the cost utility of cellular users, each cellular user needs to adaptively update the market interference price set to the user to the D2D based on the transmit power response of the user  $\tilde{P}_m^*$  (or  $\mathbf{P}_m^*$ ) by D2D. The subgame problem of the leader in the optimization problem (20) can be decomposed into two subproblems: by correcting the price factor  $c_k$ , the optimal transmit power of each cellular user  $k$  can be obtained, and then the optimal  $c_k$  is obtained. The solution of the optimal transmit power of the cellular user  $k$  can be obtained by the Lagrangian dual decomposition method.

Therefore, for the subgame problem of the leader in (20), the corresponding Lagrangian function can be obtained:

$$L_k(\tilde{p}_k, \tilde{p}_{-k}, c_k, \nu_k) = e^{\tilde{p}_k} - c_k \sum_{m=1}^M \left( e^{\tilde{p}_m^k} L_{mk} |h_{mk}|^2 \right) + \nu_k \left( \frac{\tilde{\gamma}_k N_0}{e^{\tilde{p}_k} L_k} + \sum_{m=1}^M \left( \log \left( \frac{\tilde{\gamma}_k (1 + \alpha_m) e^{\tilde{p}_m^k} L_{mk}}{e^{\tilde{p}_k} L_k} \right) \right) - \log \left( \frac{1}{1 - \xi_k} \right) \right), \tag{25}$$

where the dual variable  $\nu_k$  represents the constraint in (20); the optimal solution for the Lagrangian dual function is  $\min L_k(\tilde{p}_k, \tilde{p}_{-k}, c_k, \nu_k)$ .

$L_k(\tilde{p}_k, \tilde{p}_{-k}, \mathbf{c}, \lambda_m, \omega_m)$  is a concave function for  $\nu_k$ . Similarly, we can use the gradient method to get the optimal dual variable:

$$\begin{aligned}
\nu_k^* &= \nu_k^* + \theta \left( \frac{\tilde{\gamma}_k N_0}{e^{\tilde{p}_k} L_k} + \sum_{m=1}^M \left( \log \left( \frac{\tilde{\gamma}_k (1 + \alpha_m) e^{\tilde{p}_m^k} L_{mk}}{e^{\tilde{p}_k} L_k} \right) \right) - \log \left( \frac{1}{1 - \xi_k} \right) \right)^+ \\
&- \log \left( \frac{1}{1 - \xi_k} \right) \Big)^+,
\end{aligned} \tag{26}$$

where  $\theta$  is a sufficiently large positive real number. The iteration in (26) will converge to the optimal  $v_k^*$ . Thus, we can derive the optimal solution ( $\tilde{p}_k^*$ ) of the problem in (20) by solving the following problem. The solution of the problem is as follows.

Then, we can get the transmit power  $p_m^{k*} = \exp(\tilde{p}_m^{k*})$  of the cellular user  $k$  by the exponential change of the formula  $\tilde{p}_k^*$ .

Based on the analysis of the optimal transmit power of D2D to the user subgame and the cellular user subgame, it is discussed how the cellular user sets the optimal price according to the transmit power policy response of all D2D pairs.

If the  $m$ -th D2D interference price to the user is greater than the payment threshold, the D2D pair of users will stop renting the channel of the  $k$ -th cellular user. In other words, for each D2D pair user, if it is transmitted on the channel of the  $k$ -th cellular user, its cost utility cannot be greater than the maximum cost threshold  $U_m^{k, \max}(e^{p_k})$ . Thus, the cost for the  $m$ -th D2D pair of users should satisfy the constraint:

$$(1 + \alpha_m)e^{\tilde{p}_m^{k*}} + c_k(1 + \alpha_m)e^{\tilde{p}_m^{k*}} L_{mk} |h_{mk}|^2 \leq U_m^{k, \max}(e^{\tilde{p}_m^{k*}}). \quad (27)$$

Therefore, we can get the constraint of the price  $c_k$ , expressed as follows:

$$c_k \leq Z_m = \frac{U_m^{k, \max}(e^{\tilde{p}_m^{k*}}) - (1 + \alpha_m)e^{\tilde{p}_m^{k*}}}{(1 + \alpha_m)e^{\tilde{p}_m^{k*}} L_{mk} |h_{mk}|^2}. \quad (28)$$

As can be seen from (28), if the cellular subscriber wants the  $m$ -th D2D pair to communicate on its channel, the price set by the cellular subscriber  $k$  should not be higher than the upper limit price in (28). Therefore, the optimal interference price  $c_k^*$  is  $c_k^* = Z_m$ .

According to the above analysis, the  $k$ -th cellular user will minimize  $Q_k(e^{p_k})$  by paying the threshold vector  $\mathbf{Z} = (Z_1, Z_2, \dots, Z_M)$  and then select the appropriate overall price  $c_k^*$ .

In order to achieve Nash equalization and to converge to the optimal transmit power, we propose the following iterative power control algorithm.

- (i) Step 1: let  $t = 0$ ; for all cellular users  $k$ , initialize its transmit power  $p_k(t)$  and price  $c_k(t)$ ; for all D2D pairs, initialize  $\mathbf{P}_m(t)$
- (ii) Step 2: for each D2D user, calculate the price vector paid by the  $m$ -th D2D to all cellular users
- (iii) Step 3: update  $\lambda_k^*$  and  $\omega_k^*$  according to formula (23), respectively
- (iv) Step 4: for a given cellular user  $p_k(t)$  and  $c_k(t)$ , each D2D pair user  $m$  is solved according to equation (24) to obtain  $\mathbf{P}_m^*(t+1)$ . If  $c_k(t) > Z_k(t)$ , D2D terminates the channel using cellular user  $k$  for user  $m$
- (v) Step 5: for each cellular user  $k$ , update  $v_k^*$  according to formula (18)

- (vi) Step 6: at a given D2D user transmit power  $\mathbf{P}_m^*(t+1)$ , the cellular user  $k$  updates its transmit power  $p_m^*(t+1)$  by solving equation (19)
- (vii) Step 7: according to formulas (28) and  $\mathbf{Z}$ , the cellular user updates its price  $c_k^*(t+1)$
- (viii) Step 9: let  $t = t + 1$  repeat steps 2 to 8 until convergence

## 6. Simulation Results and Performance Analysis

This paper uses simulation to analyze the performance of the Stackelberg game power control. Assuming the radius of the cell is 500 meters, the maximum transmit power of the cellular user and D2D to the user is  $P_m^{\max} = 24$  dBm and  $P_m^{\max} = 20$  dBm, the noise power  $N_0 = 10^{-10}$  w, and the maximum distance of each cellular user from the base station is 300 meters, and the maximum communication distance of each D2D pair is 50 meters. For each D2D pair user, we set  $\alpha_m = -1$ ; self-interference relief coefficient  $\eta_1 = \eta_2 = -100$  dB.

Figure 2 shows the interrupt performance of three cellular users and five D2D pairs at different target SINR values. The outage probability of the two users becomes larger as the target SINR threshold increases. However, the terminal probability of the proposed game scheme is significantly lower than the other two power control schemes, especially at the low SINR threshold. This is because the proposed scheme takes into account the user's minimum SINR requirement and builds an optimization problem under the outage probability, which effectively reduces the probability of interruption of users in the network. The other two schemes do not consider the outage probability into the system optimization problem, so the interrupt performance is lower than the proposed scheme.

Figure 3 is a comparison of the three algorithms for cellular users and D2D throughput performance for users. As can be seen from Figure 4, as the signal-to-noise ratio increases, the total throughput of the user also increases. When the signal-to-noise ratio is low, the three algorithms can get better performance, and the performance of the proposed algorithm in this case is almost the same as other algorithms mainly because of the constraints. However, as the signal-to-noise ratio increases, the performance of the proposed scheme is improved and higher than the other two schemes. In addition, it can be seen that the performance of D2D to the user's throughput performance is not significantly improved in the case of high signal-to-noise ratio. This is because the interference power to the cellular user is less than a certain value, so the transmission power of the D2D to the user is limited. However, compared with the other two schemes, D2D still has higher throughput for users.

Figure 4 compares the three algorithms for the total cost utility of cellular users and D2D to users. As can be seen from Figure 4, as the background noise increases, the total cost utility of the user also increases. This is because, in the case of large background noise, the user only increases his own transmission power to improve the signal-

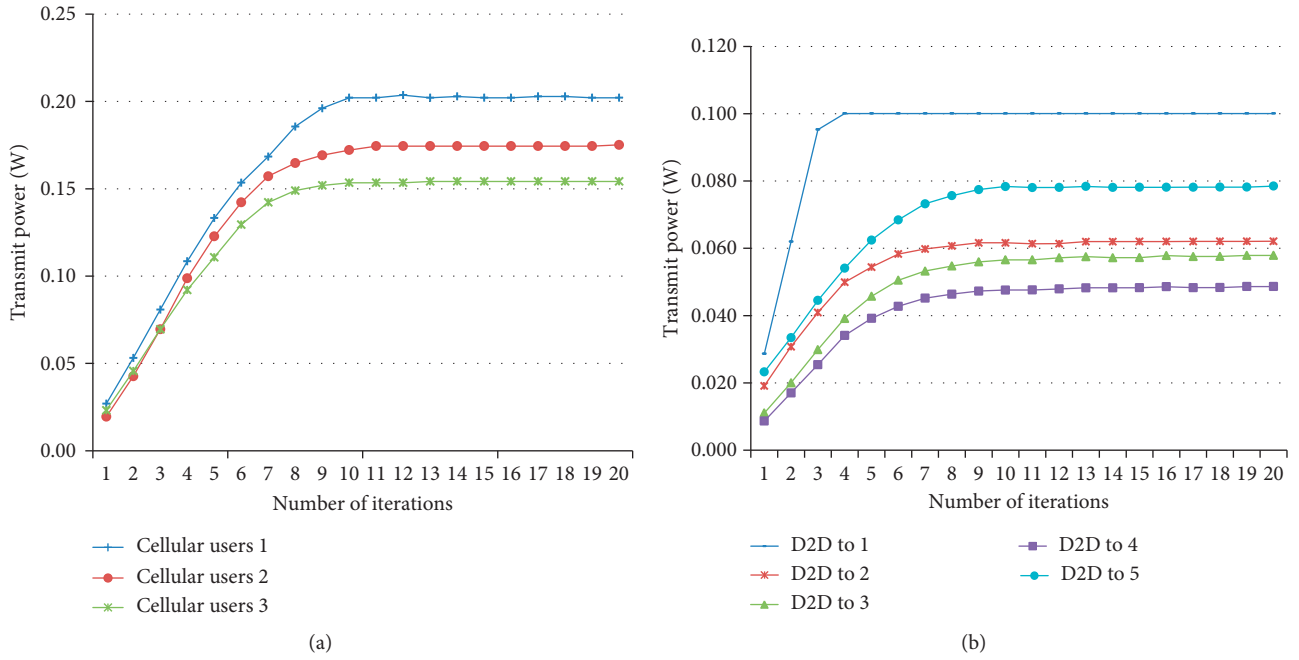


FIGURE 2: Convergence of cellular users (a) and D2D to users (b).

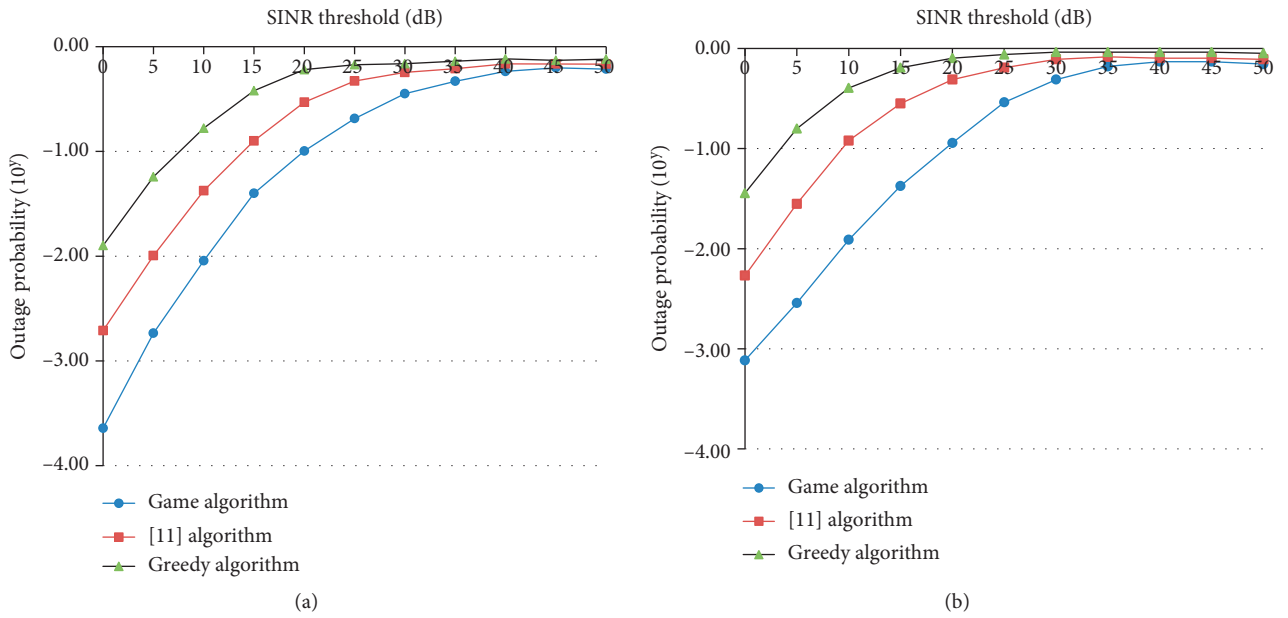


FIGURE 3: Cellular user (a) and D2D outage probability to user (b).

to-interference ratio of the receiving end to meet the requirement of the outage probability, and the cost increases. The algorithm proposed in this paper can reduce

the power cost of the network to a certain extent compared with other algorithms. In addition, it can be seen that the cost required by the greedy algorithm is the

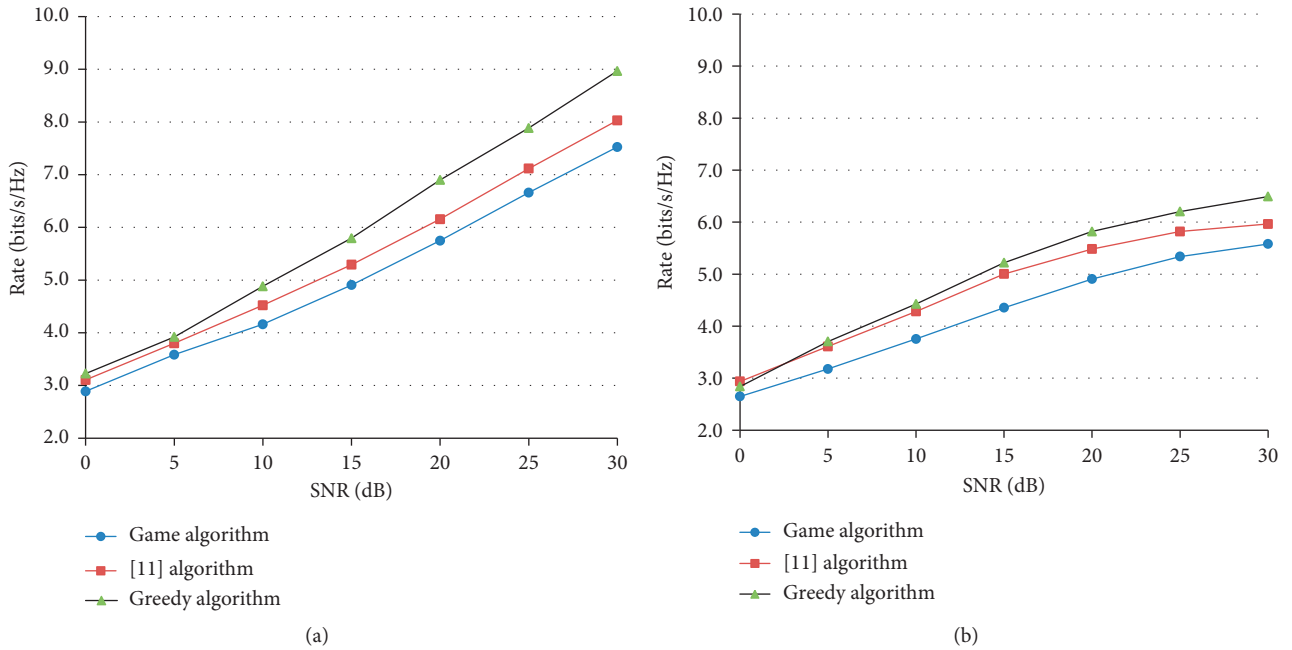


FIGURE 4: Cellular user and D2D throughput to users. (a) Cellular user. (b) D2D user.

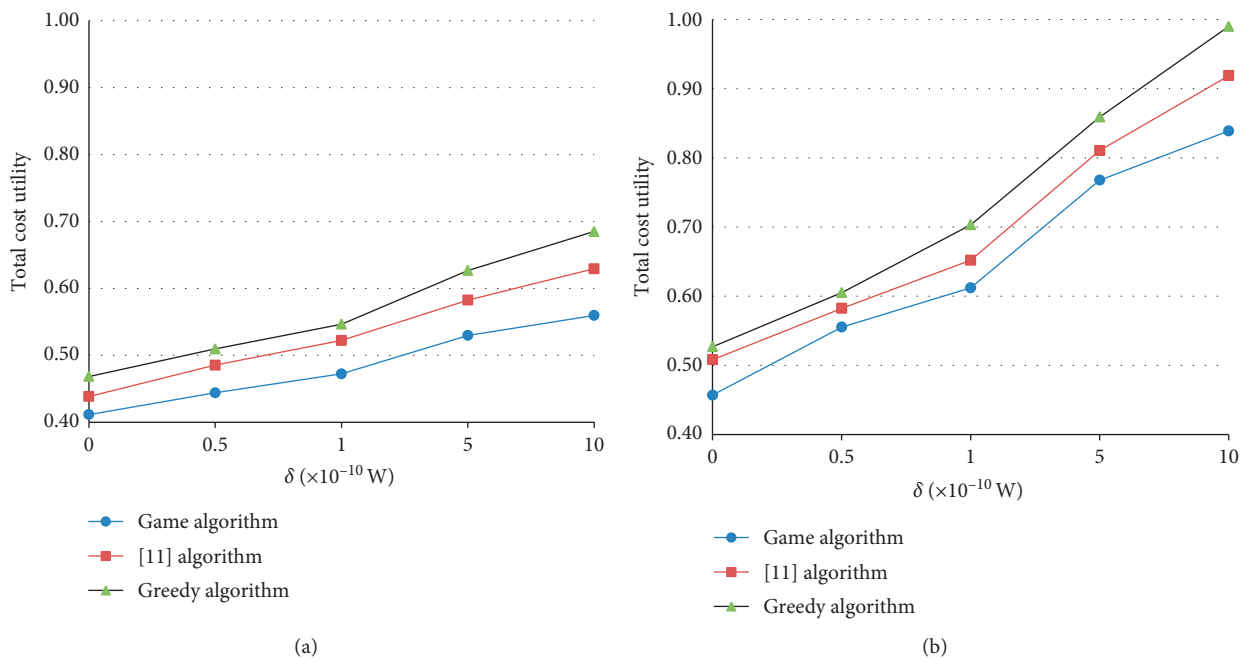


FIGURE 5: Total cost utility of cellular users and D2D to users. (a) Cellular user. (b) D2D user.

largest, because the algorithm causes the user to self-increase the transmission power to satisfy the performance in the network, and ignores the interference caused to other users. The details are shown in Figure 5.

## 7. Conclusions

This paper studies the problem of uplink power control in a full-duplex D2D communication network under interrupt

constraints. In order to effectively control the users transmit power, this paper introduces the Stackelberg game power control algorithm, setting up cellular users and D2D users as leaders and followers, respectively. Cellular users and D2D users compete with each other to minimize the cost utility and a new utility function is proposed. Under the condition of channel interruption constraint and power constraint, the optimization problem is constructed and the Lagrange dual decomposition method is used to obtain the user's



transmitting power. Finally, the performance of the proposed game scheme is analyzed and the Nash equilibrium can be converging. Future work will focus on the construction of intelligent algorithm research for D2D competition in the case of multi-users and assigning solutions through user weights.

## Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

## Acknowledgments

This work was supported by the Key Project of Science and Technology Research of Chongqing Education Commission (Grant No. KJZD-K201902101), the Open Fund of Chongqing Key Laboratory of Spatial Data Mining and Big Data Integration for Ecology and Environment, and Humanities and Social Sciences Project of Rongzhi College of Chongqing Technology and Business University (Grant No. 20197004).

## References

- [1] A. Rinne, "Optimal power allocation for full-duplex D2D communications over wireless cellular networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1466–1479, 2017.
- [2] D. Yu and X. Xueba, "Energy-efficient power control algorithm for D2D communication," *Journal of Electronics and Information Technology*, vol. 39, no. 4, pp. 997–1001, 2017.
- [3] H. Yang and X. Xie, "A robust stackelberg game based uplink power control for device-to-device communication with channel uncertainty and outage probability constraints," *Wireless Personal Communications*, vol. 90, no. 1, pp. 1–23, 2015.
- [4] A. Celik, R. M. Radaydeh, F. S. Al-Qahtani, and M.-S. Alouini, "Resource allocation and interference management for D2D-enabled DL/UL decoupled het-nets," *IEEE Access*, vol. 5, pp. 22735–22749, 2017.
- [5] Y. Zhang, Y. Ling, and Y. Lu, "Multi-cell D2D resource allocation algorithm based on non-cooperative game," *Optical Communication Research*, vol. 54, no. 2, pp. 75–78, 2016.
- [6] X. Xie, "Rate maximization power allocation algorithm for D2D full-duplex communication in cognitive networks," *Journal of Electronics & Information Technology*, vol. 39, no. 4, pp. 1002–1006, 2017.
- [7] S. Dang, J. P. Coon, and G. Chen, "Resource allocation for full-duplex relay-assisted device-to-device multicarrier systems," *IEEE Wireless Communications Letters*, vol. 6, no. 2, pp. 166–169, 2017.
- [8] L. Yun, "Resource allocation algorithm supporting full-duplex D2D communication in cellular networks," *Microelectronics & Computer*, vol. 33, no. 11, pp. 142–146, 2016.
- [9] R. Tang, J. Zhao, H. Qu, and Z. Zhang, "Energy-efficient resource allocation for 5G full-duplex enabled device-to-device communication," in *Proceedings of the IEEE Globecom Workshops (GC Wkshps)*, IEEE Washington, DC, USA, December 2016.
- [10] J. Wang and X. Zhang, "Adaptive power control for maximizing channel capacity over full-duplex D2D Q-OFDMA ad hoc networks," in *Proceedings of the IEEE Global Communications Conference*, IEEE, San Diego, CA, USA, December 2015.
- [11] K. S. Ali, H. Elsayy, and M.-S. Alouini, "Modeling cellular networks with full-duplex D2D communication: a stochastic geometry approach," *IEEE Transactions on Communications*, vol. 64, no. 10, pp. 4409–4424, 2016.
- [12] M. Zulhasnine, C. Huang, and A. Srinivasan, "Efficient resource allocation for device-to-device communication underlying LTE Network," in *Proceedings of the 2010 IEEE 6th International Conference on Wireless and Mobile Computing, Networking and Communications*, pp. 368–375, IEEE, Niagara Falls, Canada, October 2010.
- [13] H. Min, W. Seo, J. Lee, S. Park, and D. Hong, "Reliability improvement using receive mode selection in the device-to-device uplink period underlying cellular networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 2, pp. 413–418, 2011.
- [14] G. Yu, L. Xu, D. Feng, R. Yin, G. Y. Li, and Y. Jiang, "Joint mode selection and resource allocation for device-to-device communications," *IEEE Transactions on Communications*, vol. 62, no. 11, pp. 3814–3824, 2014.
- [15] X. Xiao, X. Tao, and J. Lu, "A QoS aware power optimization scheme in OFDMA systems with integrated device-to-device (D2D) communications," in *Proceedings of the 2011 IEEE Vehicular Technology Conference (VTC Fall)*, IEEE, San Francisco, CA, USA, September 2011.
- [16] Y. Hassan, F. Hussain, S. Hossen, S. Choudhury, and M. Mahbub Alam, "Interference minimization in D2D communication underlying cellular networks," *IEEE Access*, vol. 5, pp. 22471–22484, 2017.
- [17] H.-H. Nguyen, M. Hasegawa, and W.-J. Hwang, "Distributed resource allocation for D2D communications underlay cellular networks," *IEEE Communications Letters*, vol. 20, no. 5, pp. 942–945, 2016.
- [18] D. D. Ningombam and S. Shin, "Radio resource allocation and power control scheme to mitigate interference in device-to-device communications underlying LTE-a uplink cellular networks," in *Proceedings of the 2017 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 961–963, IEEE, Jeju, South Korea, October 2017.
- [19] G. U. Jaheon, S. J. Bae, S. F. Hasan, and M. Y. Chung, "A combined power control and resource allocation scheme for D2D communication underlying an LTE-advanced system," *IEEE Transactions on Communications*, vol. 96, no. 10, pp. 2683–2692, 2013.
- [20] G. Fodor, E. Dahlman, G. Mildh et al., "Design aspects of network assisted device-to-device communications," *IEEE Communications Magazine*, vol. 50, no. 3, pp. 170–177, 2012.
- [21] K. M. Doppler, C. Wijting, C. Ribeiro, and K. Hugl, "Device-to-device communication as an underlay to LTE-advanced networks," *IEEE Communications Magazine*, vol. 47, no. 12, pp. 42–49, 2009.
- [22] H. Kim, N. Jee-hyeon, and E. Cho, "Resource allocation policy to avoid interference between cellular and D2D links in mobile networks," in *Proceedings of the International Conference on Information Networking 2014 (ICOIN2014)*, pp. 588–591, IEEE, Phuket, Thailand, February 2014.

- [23] J. C. F. Li, M. Lei, and F. Gao, "Device-to-device (D2D) communication in MU-MIMO cellular networks," in *Proceedings of the 2012 IEEE Global Communications Conference (GLOBECOM)*, pp. 3583–3587, IEEE, Anaheim, CA, USA, December 2012.
- [24] A. Tsai, L.-C. Wang, J.-H. Huang, and T. Lin, "Intelligent resource management for device-to-device (D2D) communications in heterogeneous networks," in *Proceedings of the International Symposium on Wireless Personal Multimedia Communications*, pp. 75–79, IEEE, Taipei, Taiwan, September 2012.
- [25] C. U. Seo, N. B. Mandayam, and D. J. Goodman, "Efficient power control via pricing in wireless data networks," *IEEE Transactions on Communications*, vol. 50, no. 2, pp. 291–303, 2002.
- [26] Y. Kuo, J. Yang, and J. Chen, "Efficient swarm intelligent algorithm for power control game in cognitive radio networks," *IET Communications*, vol. 7, no. 44, pp. 1089–1098, 2013.
- [27] Z. Han and K. J. R. Liu, "Non-cooperative power-control game and throughput game over wireless network," *IEEE Transactions on Communications*, vol. 59, no. 10, pp. 1625–1629, 2005.

## Research Article

# Performance Investigation of Stochastic Resonance in Three Types of Asymmetric Bistable System Driven by Trichotomous Noise

Si-Hai Zhao , Jiang-Ye Xu, Yu-Xiao Liu, Ze-Xing Zhao, and Zhong-Shun Qin

School of Mechanical Electrical and Information Engineering, China University of Mining and Technology Beijing Campus, Beijing 100083, China

Correspondence should be addressed to Si-Hai Zhao; zsh@cumtb.edu.cn

Received 16 September 2020; Revised 16 October 2020; Accepted 24 October 2020; Published 10 November 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Si-Hai Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes a new system whose potential function is with three types of asymmetric potential wells, driven by trichotomous noise. Firstly, the three types of asymmetric bistable system are described in detail, and the changes of asymmetric bistable system potential function under different asymmetric factors are analyzed. Secondly, the effect of potential function parameters, asymmetric factor  $\alpha$ , noise intensity  $D$ , and the probability of particle transition  $q$  is discussed, using numerical simulation. The detection effects of traditional symmetric SR and three types of asymmetric SR are observed and compared under the driving of trichotomous noise and periodic signals. The mean of signal-to-noise ratio gain is the indicator of the system's effectiveness on enhancing weak signal. The results indicate that it can make the detection effect of the asymmetric system better than that of the traditional bistable system by adjusting the parameters of the asymmetric stochastic resonance system and trichotomous noise.

## 1. Introduction

Asymmetry is a ubiquitous phenomenon which has been a widespread concern for researchers for a long time due to its potential applications in physical and natural systems, such as neural networks [1, 2], wireless communications [3], optical metamaterials [4], and medical imaging [5]. In addition, as demonstrated in a number of studies, the effect of stochastic resonance can be enhanced in the case that the potential function is not symmetric [6–10]. However, there are relatively few studies devoted to the asymmetric system SR driven by trichotomous noise.

It is generally accepted that noise has a negative effect on the extraction and detection of weak signals. However, if the noise is used reasonably, the weak signal can be enhanced according to the stochastic resonance theory. Stochastic resonance is regarded as a method of weak signal detection [11–14] since it was firstly proposed to express the periodic alternations between the Earth's ice age and the warm

climate period in 1981 [15]. Later, various potential function models of stochastic resonance were proposed based on the model of traditional bistable stochastic resonance [16, 17], such as monostable models [18–20], tristable models [21–23], and two metastable models [24]. Then, the underdamping SR system [25–27] that is different from the traditional systems (overdamped system [28–30]) has been paid much attention by researchers. In addition, in the research of SR, the driving noise is various, including Gaussian white noise [31–33] and colored noise [34–37]. And it can be divided into multiplicative noise and additive noise [32, 38–40], according to the action mode of driving noise to the system.

Much work so far focusing on stochastic resonance always takes the Gaussian white noise as the driving noise. However, it should be noted that most of the noise in a physical system is not gaussian white noise. Trichotomous noise is a function of the time variable  $t$  that is converted between the three values of  $a$ ,  $b$ ,  $c$ . Considerable research

efforts have been devoted to SR phenomenon driven by trichotomous noise [41–44]. However, SR systems with asymmetric potential function driven by trichotomous noise have not been realized yet. Hence, the combination of trichotomous noise and asymmetric SR system presents a wide range of possibilities for the further development of weak signal detection.

## 2. Models and Methods

The Langevin equation of asymmetric stochastic resonance system is considered as follows:

$$\frac{\partial x(t)}{\partial t} = -\frac{\partial U(x)}{\partial x} + A \cos(2\pi ft) + \xi(t), \quad (1)$$

where  $A \cos(2\pi ft)$  is the system input signal to be measured,  $A$  is the amplitude of the input signal to be measured,  $f$  is the frequency of the input signal to be measured,  $\xi(t)$  is the trichotomous noise, and  $U(x)$  is the potential function of asymmetric systems.  $U(x)$  is given as follows:

$$U_i(x) = \begin{cases} \frac{a_0 x^2}{2} + \frac{b_0 x^4}{4}, & x \geq 0, \\ \frac{a_0 A_i x^2}{2} + \frac{b_0 B_i x^4}{4}, & x < 0, \end{cases} \quad (2)$$

where  $m, n, A_i,$  and  $B_i$  are the structural parameters of asymmetric systems, and these parameters meet such inequalities:  $a_0, b_0, A_i, B_i > 0$  and  $i = 1, 2, 3$ .

### 2.1. Three Kinds of Asymmetric Stochastic Resonance Systems

**2.1.1. The Bistable System with Asymmetric Depth of Potential Well.** When  $A_1 = B_1 = \alpha$ , ( $\alpha$  is asymmetric factor) and there is no input signal, the potential function contains two minima in the potential well:  $x_{\min\pm} = \pm\sqrt{(a_0/b_0)}$  (stable points) and maxima in the potential barrier  $x_{\max} = 0$  (unstable point). The widths of left and right potential well are both  $\sqrt{(a_0/b_0)}$ . Then, the depths of the wells can be obtained as follows:

$$\Delta U_L = \frac{\alpha a_0^2}{4b_0}, \quad (3)$$

$$\Delta U_R = \frac{a_0^2}{4b_0}. \quad (4)$$

As can be seen from formulae (3) and (4), only the depth of the left well is controlled by the asymmetric factor  $\alpha$ . Figure 1(a) shows the different depths of potential function caused by  $\alpha$  when  $a_0 = 1$  and  $b_0 = 1$ , where the depths of the left potential well increase with  $\alpha$  increasing.

**2.1.2. The Bistable System with Asymmetric Width of Potential Well.** When  $A_2 = (1/\alpha^2)$ ,  $B_2 = (1/\alpha^4)$ , the potential function contains two stable points,  $x_{\min+} = \sqrt{(a_0/b_0)}$  and  $x_{\min-} = -\alpha\sqrt{(a_0/b_0)}$ , and an unstable point  $x_{\max} = 0$ . The

depths of left and right potential well are both  $(a_0^2/4b_0)$ . Then, the widths of the wells can be obtained as follows:

$$\Delta W_L = \alpha\sqrt{\frac{a_0}{b_0}}, \quad (5)$$

$$\Delta W_R = \sqrt{\frac{a_0}{b_0}}. \quad (6)$$

As can be seen from formulae (5) and (6), the width of the left well is controlled by the asymmetric factor  $\alpha$ . Figure 1(b) shows the different widths of the potential function caused by  $\alpha$  when  $a_0 = 1$  and  $b_0 = 1$ , where the widths of the left potential well increase with  $\alpha$  increasing.

**2.1.3. The Bistable System with Both Asymmetric Width and Asymmetric Depth of Potential Well.** When  $A_3 \neq B_3, A_3^2 \neq B_3$ , the bistable system has potential wells with both asymmetric width and depth. In this research,  $A_3 = 1, B_3 = (1/\alpha^4)$  (the combination of  $U_1(x)$  and  $U_2(x)$ ) are selected as typical and the potential function contains two stable points,  $x_{\min+} = \sqrt{(a_0/b_0)}$  and  $x_{\min-} = -\alpha^2\sqrt{(a_0/b_0)}$ , and an unstable point  $x_{\max} = 0$ . Then, the depths and widths of the wells can be obtained as follows:

$$\Delta U_L = \frac{\alpha^4 a_0^2}{4b_0}, \quad (7)$$

$$\Delta U_R = \frac{a_0^2}{4b_0}, \quad (8)$$

$$\Delta W_L = \alpha^2\sqrt{\frac{a_0}{b_0}}, \quad (9)$$

$$\Delta W_R = \sqrt{\frac{a_0}{b_0}}. \quad (10)$$

As can be seen from formulae (7)–(10), the left well is controlled by the asymmetric factor  $\alpha$ . Figure 1(c) shows the differences of the potential function caused by  $\alpha$  when  $a_0 = 1$  and  $b_0 = 1$ , where the widths and depths of the left potential well increase with  $\alpha$  increasing.

**2.2. Method of Generating Trichotomous Noise.** In (1),  $\xi(t)$  is trichotomous noise, and the method of generating it will be introduced in this section. Trichotomous noise belongs to colored noise. The transitions state of trichotomous noise is between three values ( $a, b,$  and  $c$ ) [42] and it is divided into symmetry trichotomous noise ( $a = -c, b = 0$ ) and asymmetry trichotomous noise ( $a \neq -c, b \neq 0$ ) according to the fluctuation values. The trichotomous process is a Poisson process, and the jump between the three values occurs with the constant probabilities:

$$\begin{aligned} P_s(a) &= P_s(c) = q, \\ P_s(b) &= 1 - 2q. \end{aligned} \quad (11)$$

The transition probabilities between the three states can be obtained as follows:

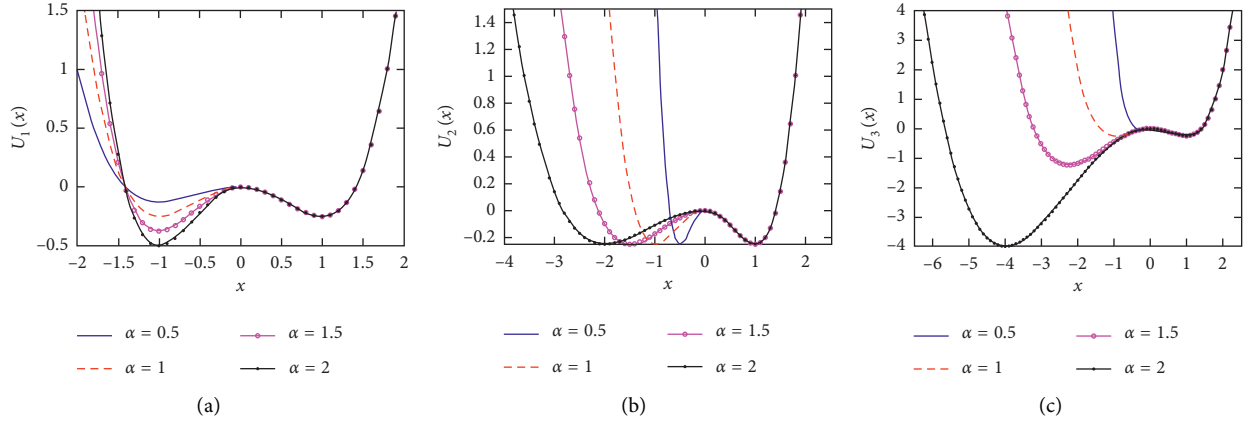


FIGURE 1: (a) Potential  $U_1(x)$  for different values of  $\alpha$ . (b) Potential  $U_2(x)$  for different values of  $\alpha$ . (c) Potential  $U_3(x)$  for different values of  $\alpha$ .

$$\begin{aligned}
 P_{aa} &= P(a, t + \tau | a, t) = q - (q - 1)e^{-\nu\tau}, \\
 P_{ba} &= P_{ca} = P(a, t + \tau | b, t) = P(a, t + \tau | c, t) = q(1 - e^{-\nu\tau}), \\
 P_{bb} &= P(b, t + \tau | b, t) = 1 - 2q(1 - e^{-\nu\tau}), \\
 P_{ab} &= P_{cb} = P(b, t + \tau | a, t) = P(b, t + \tau | c, t) = (1 - 2q)(1 - e^{-\nu\tau}), \\
 P_{cc} &= P(c, t + \tau | c, t) = q - (q - 1)e^{-\nu\tau}, \\
 P_{ac} &= P_{bc} = P(c, t + \tau | a, t) = P(c, t + \tau | b, t) = q(1 - e^{-\nu\tau}).
 \end{aligned} \tag{12}$$

Here,  $\tau > 0$ ,  $0 < q < (1/2)$ ,  $\nu > 0$ ,  $\tau > 0$ ,  $\nu = (1/\tau_{\text{cor}})$ , and  $\tau_{\text{cor}}$  is the noise correlation time. The trichotomous process is completely determined by (11) and (12). Thus, the mean value  $\langle \xi(t) \rangle$  and correlation function  $\langle \xi(t + \tau)\xi(t) \rangle$  of the trichotomous noise in the steady state can be calculated as

$$\langle \xi(t) \rangle = aq + b(1 - 2q) + cq, \tag{13}$$

$$\langle \xi(t + \tau)\xi(t) \rangle = \langle \xi(\tau) \rangle_s^2 + \frac{D\nu}{2}e^{-\nu\tau}. \tag{14}$$

Here, the noise intensity  $D$  is defined as

$$\begin{aligned}
 D &= \int_0^\infty (\langle \xi(\tau)\xi(0) \rangle_s - \langle \xi(\tau) \rangle_s^2) d\tau, \\
 &= \frac{2}{\nu} (a^2q + b^2(1 - 2q) + c^2q - [aq + b(1 - 2q) + cq]^2).
 \end{aligned} \tag{15}$$

According to the description above, the numerical method is used to generate trichotomous noise. Firstly, the function  $\xi(t)$  starts at  $a$ ,  $\xi(0) = a$ . Then, The uniformly distributed random function  $R(t)$  with value range  $[0, 1]$  is used to compare with the conditional probability given in (12) in order to determine the value of  $\xi(0 + \Delta t)$ . If  $R(0) < P_{aa}$ , the value of function  $\xi(0 + \Delta t)$  is  $a$ , and if  $P_{aa} < R(0) < P_{aa} + P_{ab}$ , the value of function  $\xi(0 + \Delta t)$  is  $b$ ; otherwise, the value of function  $\xi(0 + \Delta t)$  is  $c$ . And then, if the value of function  $\xi(0 + \Delta t)$  is  $b$ ,  $R(0 + \Delta t)$  will be used to determine the value of  $\xi(0 + 2\Delta t)$ . If  $R(0 + \Delta t) < P_{ba}$ , the value of function  $\xi(0 + 2\Delta t)$  is  $a$ , and if

$P_{ba} < R(0) < P_{ba} + P_{ba}$ , the value of function  $\xi(0 + 2\Delta t)$  is  $b$ ; otherwise, the value of function  $\xi(0 + 2\Delta t)$  is  $c$ . The case where the value of  $\xi(0 + \Delta t)$  is  $c$  is similar to the case where the value of  $\xi(0 + \Delta t)$  is  $b$ . By repeating the procedure, we can obtain the trichotomous noise  $\xi(t)$ .

Figures 2(a) and 2(b) show the profiles of the asymmetric trichotomous noise with parameters  $a = 3$ ,  $b = 1$ ,  $c = -2$ . Figures 2(c) and 2(d) show the profiles of the symmetric trichotomous noise with parameters  $a = 2$ ,  $b = 0$ ,  $c = -2$ . By comparing Figures 2(a) and 2(b), it is considered that the residence time extends with the increase of noise intensity  $D$ . When  $q = (1/2)$ , it shows that the trichotomous noise reduces to dichotomous noise in Figure 2(d).

### 3. Performance Index

There are many kinds of indexes to measure stochastic resonance, such as signal-to-noise ratio (SNR) [45, 46] and signal-to-noise ratio gain (SNRG) [47], and mean first-passage time (MFPT) [46]. In practical engineering, people are more concerned about the gain of the output SNR compared with the input SNR, so the SNR gain is chosen in this paper, given as

$$\text{SNRG} = \frac{(S_{\text{out}}(f_0)/\xi_{\text{out}}(f_0))}{(S_{\text{in}}(f_0)/\xi_{\text{in}}(f_0))}, \tag{16}$$

where  $S_{\text{out}}(f_0)$  and  $S_{\text{in}}(f_0)$  represent the output and input power of the signal at input frequency  $f_0$ , and  $\xi_{\text{out}}(f_0)$  and  $\xi_{\text{in}}(f_0)$  represent the output and input power of the noise at input frequency  $f_0$ . Due to the randomness of the noise, it is usually necessary to conduct numerous experiments and take the average value of them to ensure the reliability of the experimental results. Therefore, the average value of signal-to-noise ratio gain is defined as

$$A - \text{SNRG} = \frac{1}{n} \sum_{j=0}^{j=n} \text{SNRG}_j, \tag{17}$$

where  $n$  is the time of simulation and  $n = 20$  in this paper.  $\text{SNRG}_j$  is SNRG of the  $j$ -th simulation.

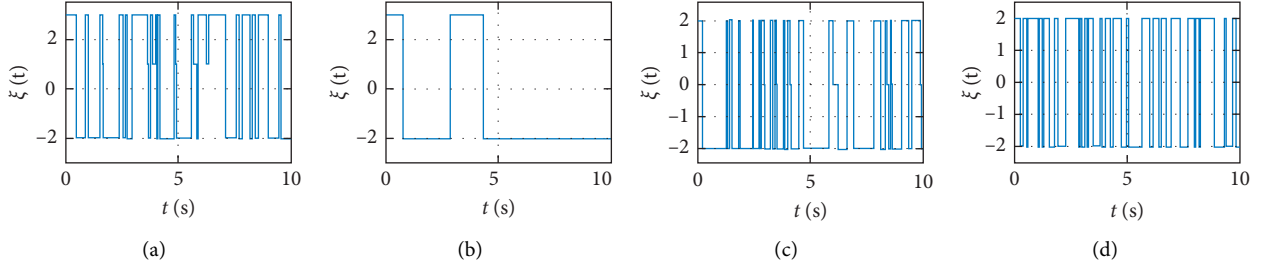


FIGURE 2: (a, b) The asymmetric trichotomous noise with parameters  $a=3$ ,  $b=1$ ,  $c=-2$ . (a)  $D=1$ ,  $q=0.3$ . (b)  $D=4$ ,  $q=0.3$ . (c, d) The symmetric trichotomous noise with parameters  $a=2$ ,  $b=0$ ,  $c=-2$ . (c)  $D=0.5$ ,  $q=0.3$ . (d)  $D=1$ ,  $q=0.5$ .

#### 4. Stochastic Resonance of Asymmetric Bistable System Driven by Trichotomous Noise

The purpose of this research is to cover the impact of system parameters and refer to the stochastic resonance of the asymmetric bistable system driven by trichotomous noise system parameters. The input sinusoidal signal is  $s(t) = A \cos(2\pi f_0 t)$ , where  $A = 0.3$ ,  $f_0 = 0.01$  Hz. The parameters of trichotomous noise distribution are set to be  $a = 1$ ,  $b = 0$ ,  $c = -1$ ,  $q = 0.3$ , and  $D = 0.5$ . A-SNRG is chosen to be the indicator. Simulation time is  $n = 50$ , for each simulation experiment, the calculation step is set as  $h = (1/f_s)$ , where the sampling frequency is  $f_s = 5$  Hz, and the sampling number is  $N = 4096$ . The optimal parameter pair  $(a, b)$  is discovered by using the adaptive algorithms in Figure 3. Figure 3 shows a three-dimensional plot of the traditional bistable adaptive optimization in which the asymmetric factor is  $\alpha = 1$ . Figure 4(a) shows the time-domain diagram of the noisy signal which contains the trichotomous noise and sinusoidal signal. It is obvious that the input signal is completely submerged by trichotomous noise. Figure 4(b) is the power spectrum diagram of the noisy signal. It could be seen from the figure that the frequency ( $f_0 = 0.01$  Hz) of the input signal cannot be identified easily.

Therefore, both passing the stochastic resonance system and the addition of noise are used to enhance the weak signal to be measured. Figure 5(a) shows the time-domain diagram of the output signal through the traditional bistable system SR system and the power spectrum of it. Figures 5(b)–5(d) show the diagram of the output signal through the bistable system with asymmetric depth, asymmetric width, and both asymmetric depth and width of potential. The peak of the asymmetric system power spectrum at  $f_0 = 0.01$  Hz is higher than the symmetric system. So, the conclusion is that the asymmetric system which contains trichotomous noise and the periodic signal has a better detection effect than the symmetric systems, by changing the depth and width of one potential well.

Next, the parameter of noise and potential function are explored under different values of the asymmetric factor in order to obtain better detection parameters.

**4.1. Influences of Asymmetric Factor  $\alpha$  for SR with Asymmetric Depth of Potential Well.** The parameters of the input signal contain  $A_0$ ,  $f_0$  and trichotomous noise contains  $a$ ,  $b$ ,  $c$ ,

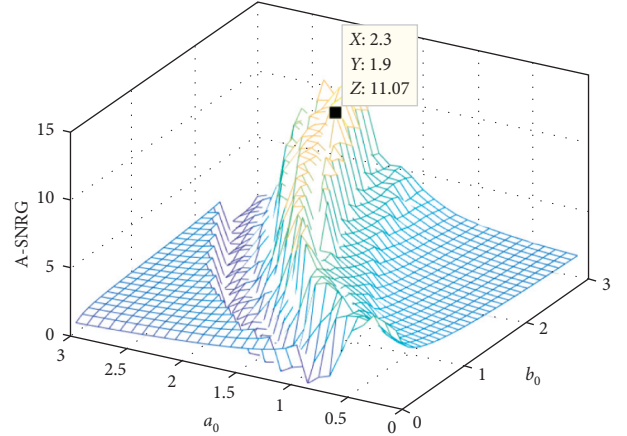


FIGURE 3: Three-dimensional diagram of the traditional symmetric bistable system SR adaptive optimization driven by trichotomous noise.

consistent with the above results. The parameters of the potential function are  $a_0 = 1.9$ ,  $b_0 = 2.3$  from the adaptive algorithms. The influence of the noise intensity  $D$  is shown in Figure 6. The influence of  $a_0$  and  $b_0$  with the noise intensity fixed  $D = 0.5$  is shown in Figure 7. Then, the influence of  $q$  with the noise intensity fixed  $D = 0.5$  and  $a_0 = 1.9$ ,  $b_0 = 2.3$  is shown in Figure 8.

Figure 6(a) shows that with the increase of the noise intensity  $D$ , the curves of A-SNRG first increase and then decrease for different values of  $\alpha$  and it means that stochastic resonance with asymmetric depth of the potential well is generated under the drive of trichotomous noise. To see more clearly, the values of the asymmetric factor are 0.6, 1.0, and 1.4 given in Figure 6(b), respectively. The maximum of A-SNRG increases with an increase in depth of the potential well, but the deeper the well, the greater the noise intensity needed to reach the peak and the reason for this is the particles need larger noise intensity  $D$  to cross the barrier. The value of A-SNRG when  $\alpha > 1$  is larger than it when  $\alpha < 1$ . Figures 6(c) and 6(d) show the SR curves for  $\alpha > 1$  and  $\alpha < 1$ , respectively. Figure 6(c) shows that the value of A-SNRG is the maximum for  $\alpha = 1.4$  and starts to decrease from 1.4. When  $\alpha = 2$ , the system cannot produce SR. Figure 6(d) shows that the curve has no peaks which means the system cannot produce SR when  $\alpha = 0.4$  and the value of A-SNRG is starting to increase. These four figures demonstrate that a

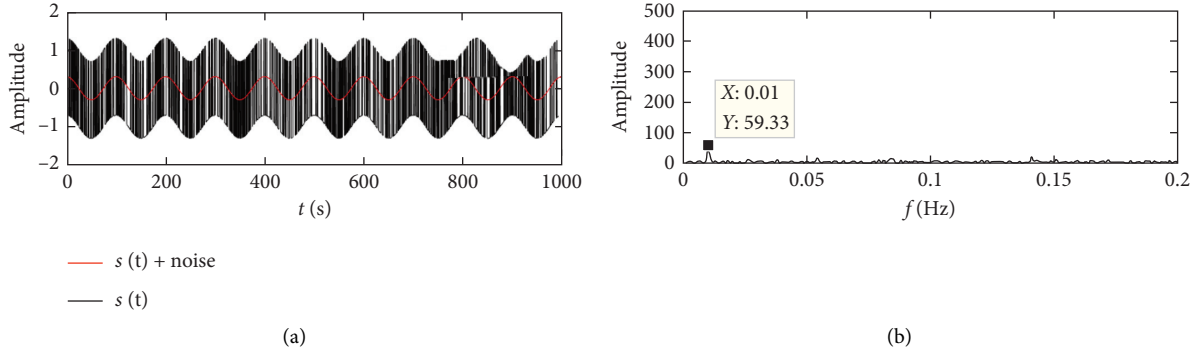


FIGURE 4: (a) The time-domain diagram of noisy signal and (b) the power spectrum diagram of the noisy signal.

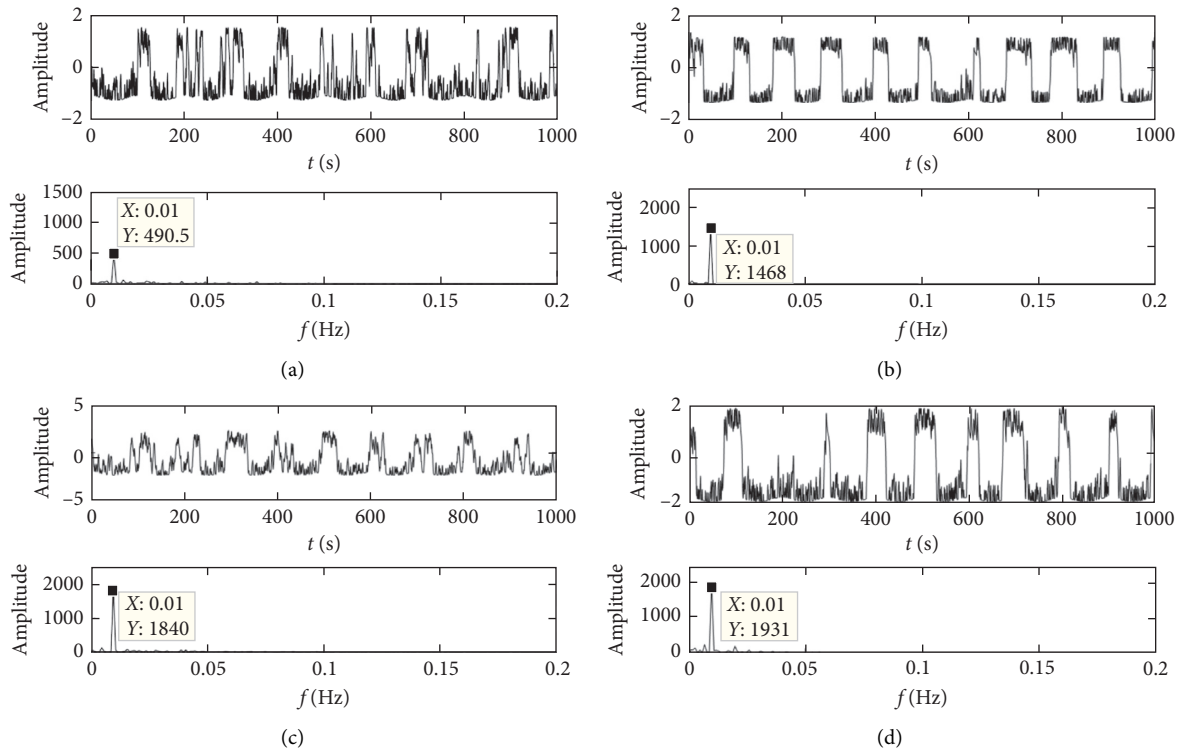


FIGURE 5: (a) Output of the symmetric system SR model and its power spectrum. (b) Output of SR model with asymmetric depth of the potential well ( $u_1(x)$ ) and its power spectrum. (c) Output of SR model with asymmetric width of potential well ( $u_2(x)$ ) and its power spectrum. (d) Output of SR model with asymmetric depth and width of potential well ( $u_3(x)$ ) and its power spectrum.

larger depth of a single potential well has a positive effect on enhancing the A-SNRG in the case that SR can be produced.

Under the condition that SR can be produced, in Figure 7(a), with the increase of parameter  $\alpha$ , the peaks of A-SNRG curves increase and the values  $a_0$  corresponding to the peak decrease. In Figure 7(b), with the increase of parameter  $\alpha$ , the peaks of A-SNRG curves increase and the values  $b_0$  corresponding to the peak increase at the same time. That is to say, there is a pair of optimal system parameters which can enhance the effect of SR.

In Figure 8, when  $\alpha < 1$ , the value of A-SNRG increases with  $q$  increasing. When  $\alpha > 1$ , the value of A-SNRG increases first and then decreases with  $q$  increasing. That is to say, for the case of shallow potential well depth, the longer

the residence time of particle is, the better the effect of stochastic resonance will be. However, when the potential well is deep, there is an optimal  $q$  for the stochastic resonance system and the value of  $q$  does not enhance the effect of the SR system significantly.

**4.2. Influences of Asymmetric Factor  $\alpha$  for SR with Asymmetric Width of Potential Well.** The parameters of the input signal contain  $A_0$ ,  $f_0$  and trichotomous noise contains  $a$ ,  $b$ ,  $c$ ,  $q$  consistent with the above results. The parameters of the potential function are  $a_0 = 1.9$ ,  $b_0 = 2.3$  from the adaptive algorithms. The influence of the noise intensity  $D$  is shown in Figure 9. Then, the influence of  $a_0$  and  $b_0$  with the noise

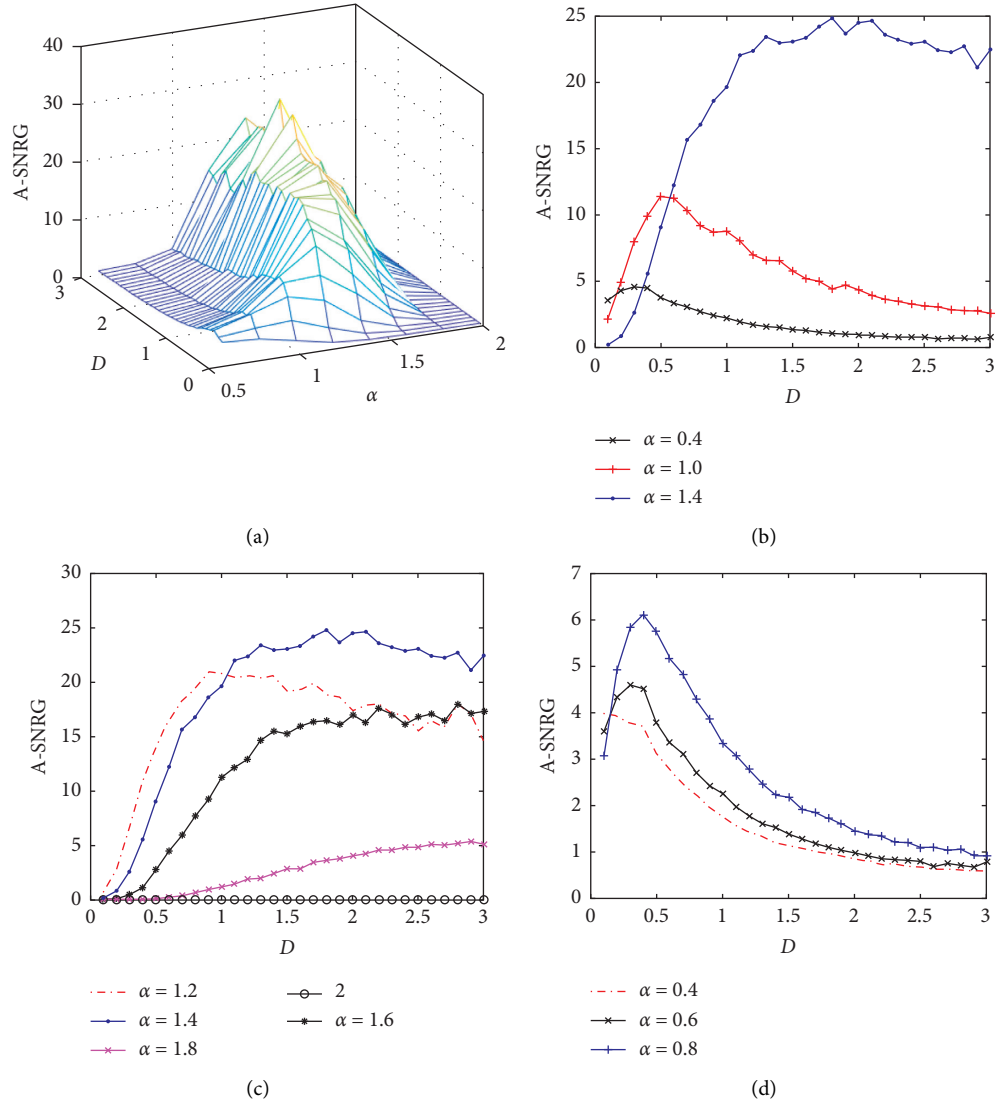


FIGURE 6: (a) A-SNRG of SR with asymmetric depth of potential well as a function of noise intensive  $D$  and asymmetric factor with different values of  $\alpha$ . (b) A-SNRG of SR with asymmetric depth of potential well as a function of noise intensive  $D$  for different values of the asymmetric factor with  $\alpha$ . (c)  $\alpha > 1$  and (d)  $\alpha < 1$ .

intensity fixed  $D = 0.5$  is shown in Figure 10. After that, the influence of  $q$  with the noise intensity fixed  $D = 0.5$ ,  $a_0 = 1.9$ ,  $b_0 = 2.3$  is shown in Figure 11.

Figure 9 shows that, with the increase of the noise intensity  $D$ , the curves of A-SNRG first increase and then decrease for different values of  $\alpha$  and it means that stochastic resonance with asymmetric width of the potential well is generated under the drive of trichotomous noise. The values of the asymmetric factor are 0.5, 0.8, 1.0, 1.2, and 1.5 given in Figure 9, respectively.

The maximum of A-SNRG decreases with an increase in width of the potential well, and the wider the well is, the greater the noise intensity will be needed to reach the peak. Compared with Figure 6(a), the changes in noise intensity which is needed to reach the peak for different depth of the potential well are larger than those of different widths. That is to say, the depth of the potential well has more effect on the

SR systems than the width. Obviously, the A-SNRG for the small  $\alpha$  is always larger than the big  $\alpha$  and the result exhibits the system with a smaller asymmetric factor  $\alpha$  which means the potential function with narrower width has a better detection effect. The value of A-SNRG when  $\alpha > 1$  is much larger than it when  $\alpha < 1$ . So the smaller width of one potential well in an asymmetric bistable system has a positive effect on stochastic resonance.

In Figure 10(a), with the increase of parameter  $\alpha$ , the peaks of A-SNRG curves increase and the values  $a_0$  corresponding to each peak decrease. And it indicates that the smaller the value of  $a_0$  is, the better the detection effect will be with a certain  $b_0$ . In Figure 10(b), with the increase of parameter  $\alpha$ , the peaks of A-SNRG curves increase and the values  $b_0$  corresponding to each peak increase at the same time. And it indicates that the greater the value of  $b_0$  is, the better the detection effect will be with a certain  $a_0$ . That is to



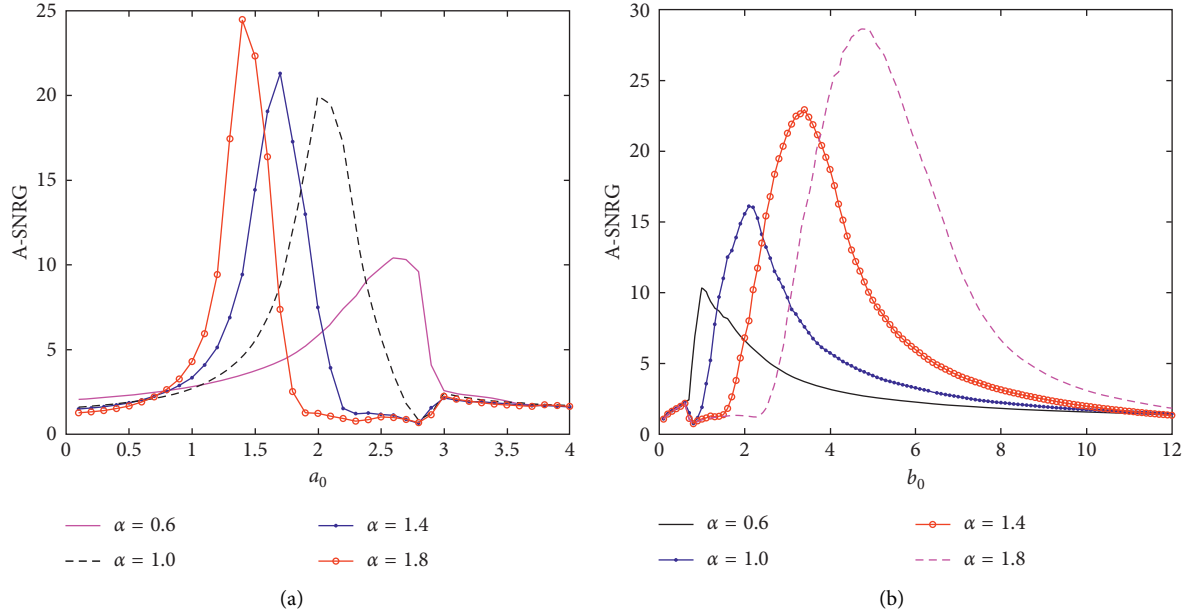


FIGURE 7: A-SNRG of SR with asymmetric depth of potential well as a function of the potential function parameter. (a)  $a_0$  and (b)  $b_0$ .

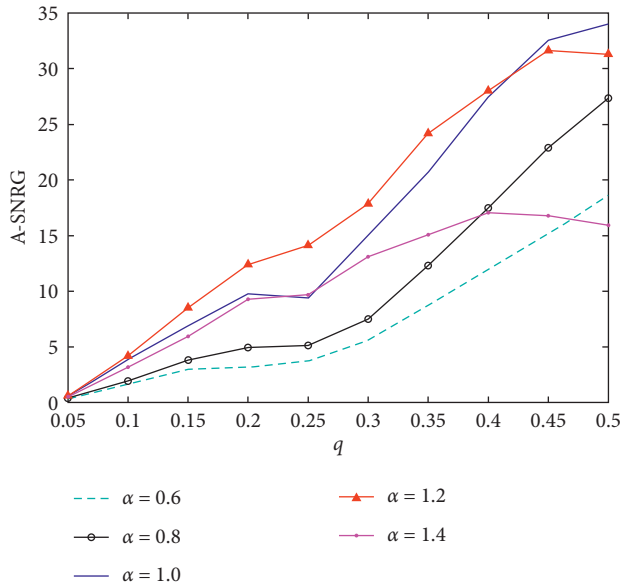


FIGURE 8: A-SNRG of SR with asymmetric depth of potential well as a function of potential function parameter  $q$ .

say, there is a pair of optimal system parameters which can enhance the effect of SR.

In Figure 11, the value of A-SNRG increases with  $q$  increasing for different values of  $\alpha$ . It means that the longer the residence time of particle is, the better the effect of the stochastic resonance will be with different widths of a single well. Compared with Figure 8, the width of the potential well has no effect on the trend of the curve for  $q$ -A-SNRG.

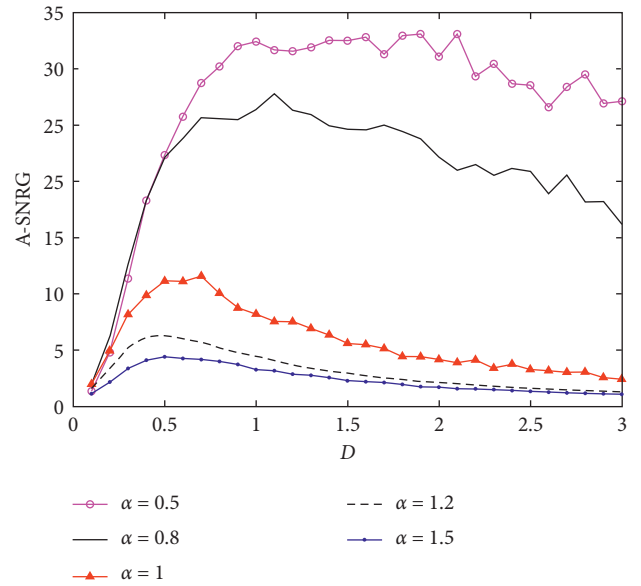


FIGURE 9: A-SNRG of SR with asymmetric width of potential well as a function of noise intensive  $D$  for asymmetric factor with different values of  $\alpha$ .

4.3. Influences of Asymmetric Factor  $\alpha$  for SR with Asymmetric Depth and Width of Potential Well. The parameters of the input signal contain  $A_0$ ,  $f_0$  and trichotomous noise contains  $a$ ,  $b$ ,  $c$  consistent with the above results. The parameters of the potential function are  $a_0 = 1.9$ ,  $b_0 = 2.3$  from the adaptive algorithms. The influence of the noise intensity  $D$  is shown in Figure 12. Then, the influence of  $a_0$  and  $b_0$  with the noise intensity fixed  $D = 0.5$  is shown in Figure 13. After

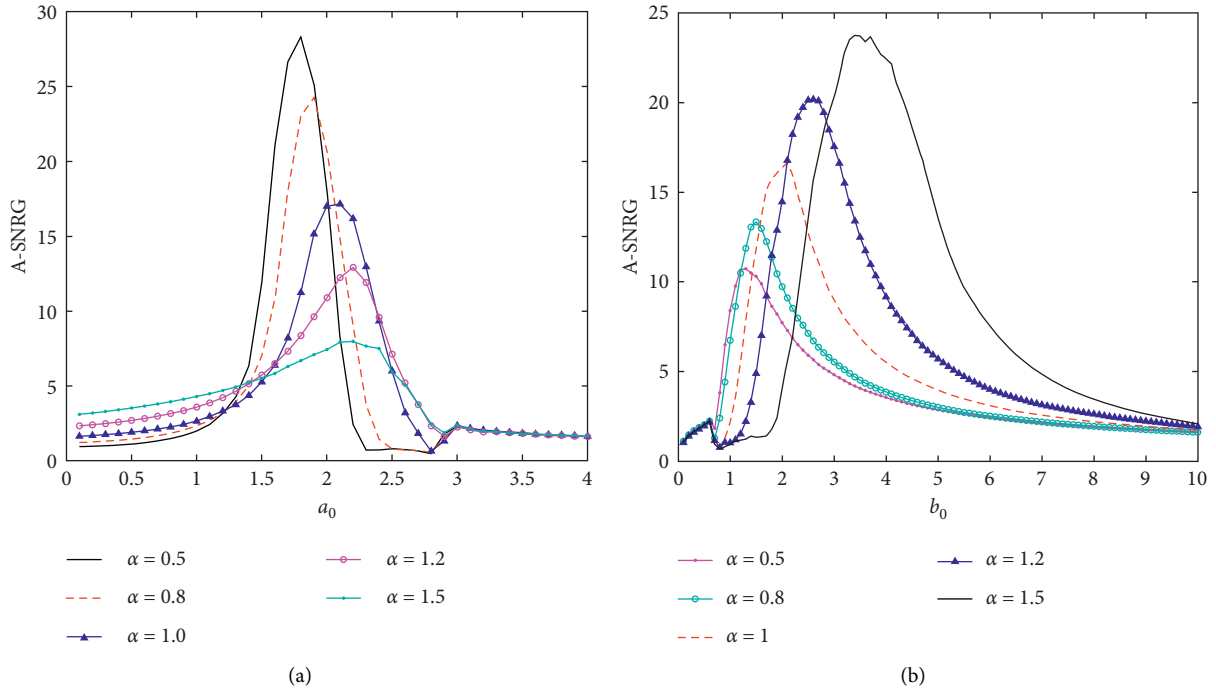


FIGURE 10: A-SNRG of SR with asymmetric width of potential well as a function of a potential function parameter. (a)  $a_0$  and (b)  $b_0$ .

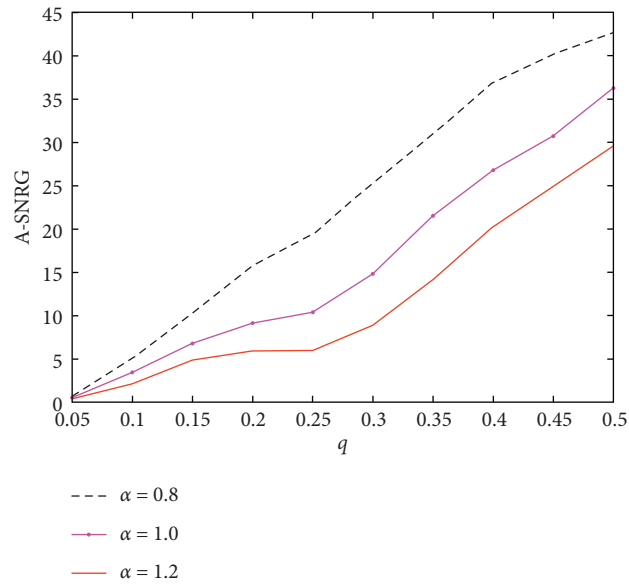


FIGURE 11: A-SNRG of SR with asymmetric width of potential well as a function of potential function parameter  $q$ .

that, the influence of  $q$  with the noise intensity fixed  $D = 0.5$  and  $a_0 = 1.9$ ,  $b_0 = 2.3$  is shown in Figure 14.

Figure 12(a) shows that when  $\alpha \in (0.8, 1.1]$ , with the increase of the noise intensity  $D$ , the curves of A-SNRG first increase and then decrease for different values of  $\alpha$  and it means that stochastic resonance with asymmetric depth and width of the potential well is generated under the drive of

trichotomous noise. To see more clearly, the values of the asymmetric factor are 0.8, 1.0, and 1.2 given in Figure 12(b), respectively. When  $\alpha = 0.8$ , the curve of A-SNRG shows a downward trend in all the values of  $D \in (0, 3]$  which means the system cannot produce SR. When  $D > 1$ , the value of A-SNRG for  $\alpha = 1.2$  is larger than that for  $\alpha = 1$ . In the case that other parameters are the same and with the appropriate

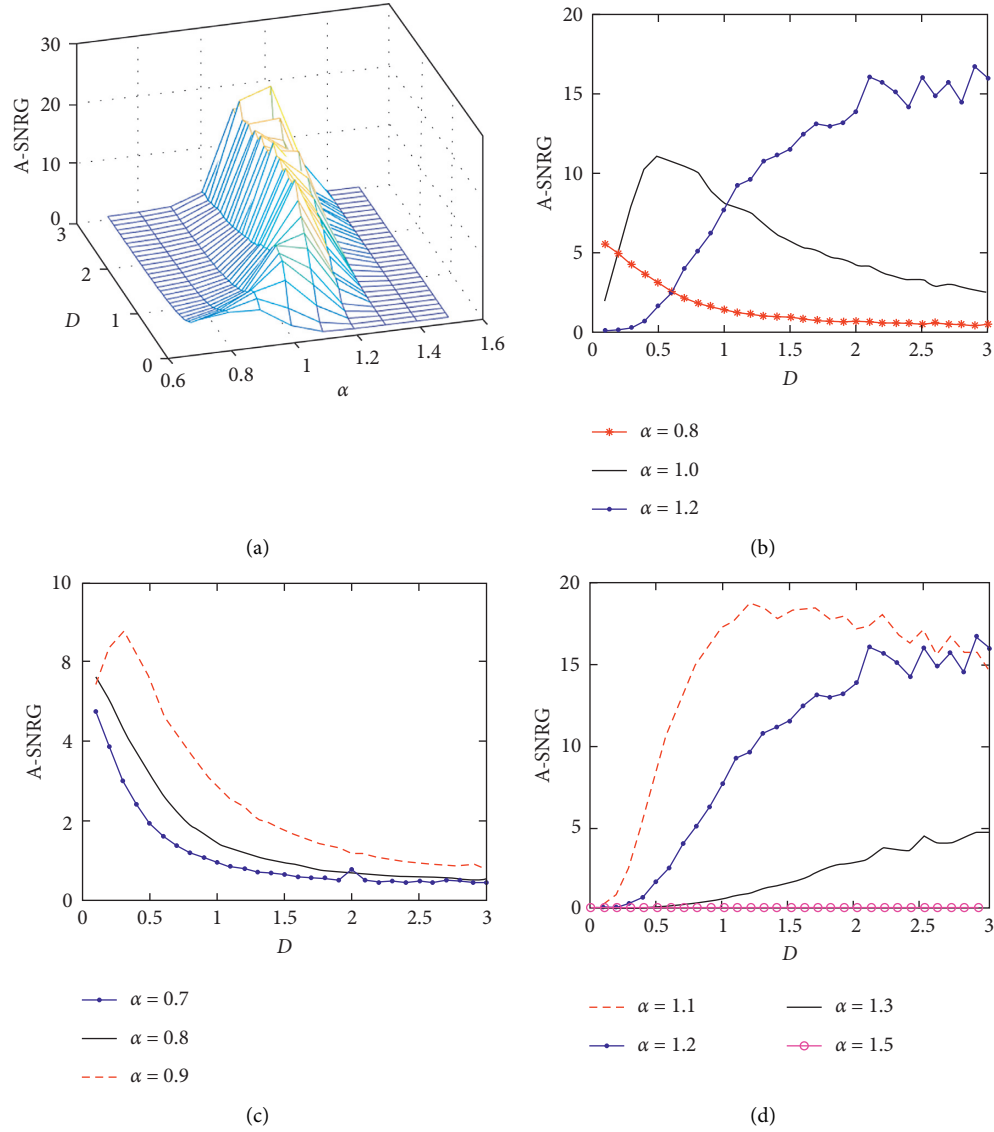


FIGURE 12: (a) A-SNRG of SR with asymmetric depth and width of potential well as a function of noise intensive  $D$  and asymmetric factor with different values of  $\alpha$ . (b) A-SNRG of SR with asymmetric depth and width of potential well as a function of noise intensive  $D$  for different values of asymmetric factor with  $\alpha$ , (c)  $\alpha < 1$  and (d)  $\alpha > 1$ .

noise intensity, the system with the asymmetric well has a more positive effect on the SR than the traditional one. The value of A-SNRG when  $\alpha > 1$  is larger than that when  $\alpha \leq 1$  under the condition that SR can be produced in Figure 12(b), and it means a larger asymmetric factor may have a better detection effect. Next, SR curves for  $\alpha < 1$  and  $\alpha > 1$  will be analyzed more clearly and Figures 12(c) and 12(d) show the SR curves for  $\alpha < 1$  and  $\alpha > 1$ , respectively. Figure 12(c) shows that the curves have no peaks which means the system cannot produce SR until  $\alpha = 0.9$ . Figure 12(d) shows that the value of A-SNRG is maximum for  $\alpha = 1.1$  and it decreases from  $\alpha = 1.1$ . When  $\alpha = 1.5$ , the values of A-SNRG are all under 1, so the system cannot produce SR with  $\alpha = 1.5$ . These four figures demonstrate that there is an

optimum asymmetric factor that has a better detection effect than the traditional system.

Under the condition that SR can be produced, in Figure 13(a), with the increase of parameter  $\alpha$ , the peaks of A-SNRG curves decrease and the values  $a_0$  corresponding to the peak decrease at the same time. In Figure 13(b), with the increase of parameter  $\alpha$ , the values  $b_0$  corresponding to the peak increase at the same time. The peaks of A-SNRG curves remain stable when  $\alpha \in [0.9, 1.1]$  and then decrease. That is to say, when  $\alpha$  is around 1,  $b_0$  is meaningless to enhance the effect of SR.

In Figure 14, when  $\alpha < 1$ , the value of A-SNRG increases with  $q$  increasing. when  $\alpha > 1$ , the value of A-SNRG increases first and then decreases with  $q$  increasing. According to

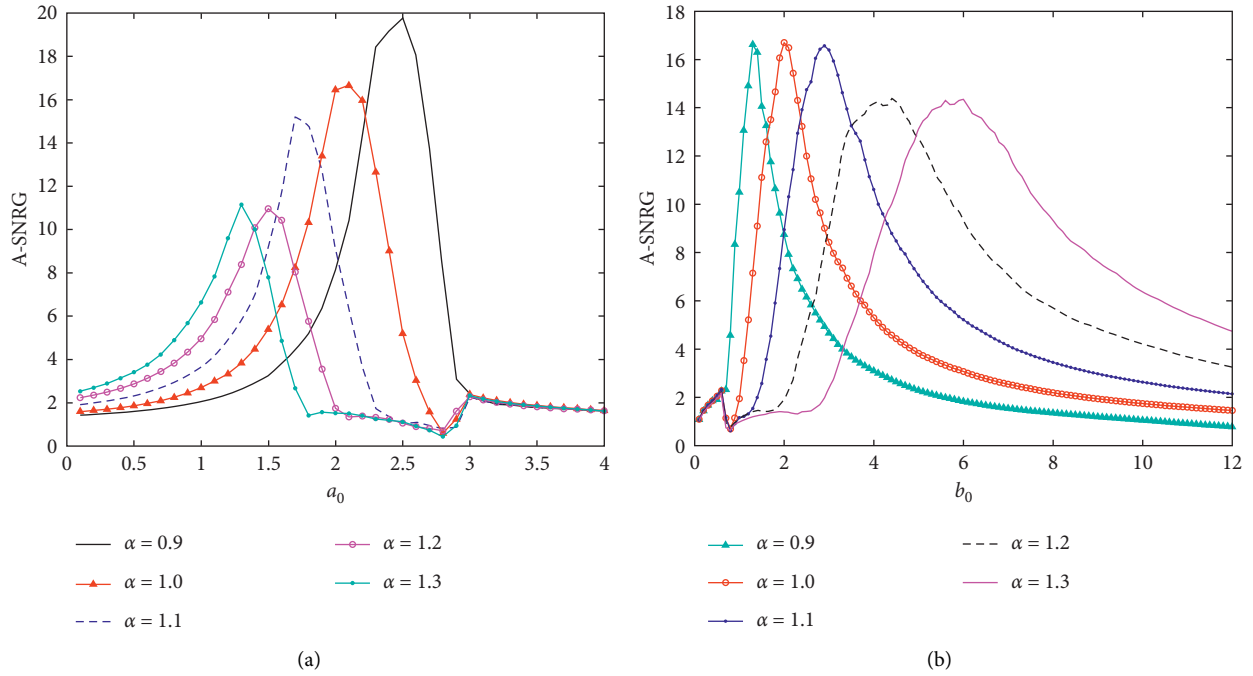


FIGURE 13: A-SNRG of SR with asymmetric depth and width of the potential well, as a function of the potential function parameter. (a)  $a_0$  and (b)  $b_0$ .

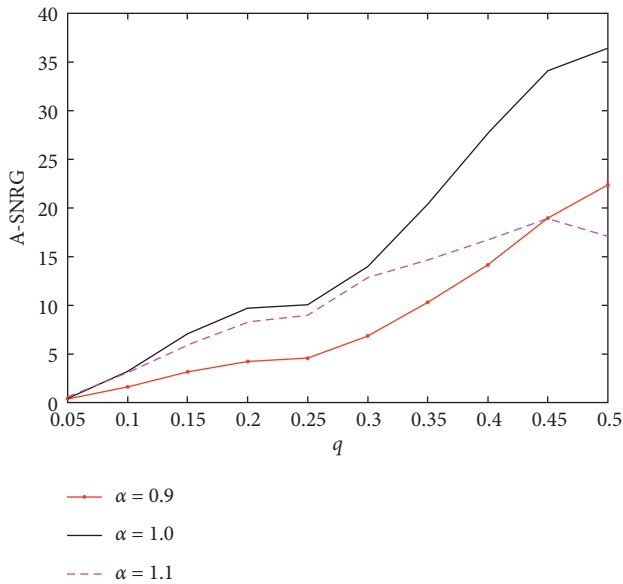


FIGURE 14: A-SNRG of SR with asymmetric depth and width of potential well as a function of potential function parameter  $q$ .

Figures 8 and 11, the trend of the curve when  $\alpha > 1$  is determined by the depth of the well only.

## 5. Conclusion

In this paper, we have focused on the SR phenomenon in three types of the asymmetric bistable system driven by trichotomous noise. The following conclusions can be drawn

from the above research: (1) when the system parameters of the potential function are different functions of an asymmetric factor, the depth and width of a single potential well can be changed. (2) When the noise intensity  $D$  is relatively large, a better detection effect can be obtained by making the depth of a single well deeper. It can improve the detection effect by making the width of the potential well smaller for different noise intensities  $D$ . (3) Parameters  $a_0, b_0$  of the potential function always have the optimal value regardless of the form of asymmetry. (4) The effect of particle residence time on the stochastic resonance system is affected by the depth of the potential well.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References


- [1] H. Sompolinsky and I. Kanter, "Temporal association in asymmetric neural networks," *Physical Review Letters*, vol. 57, no. 22, p. 2861, 1986.
- [2] K. Gopalsamy and X.-z. He, "Stability in asymmetric hopfield nets with transmission delays," *Physica D: Nonlinear Phenomena*, vol. 76, no. 4, p. 344, 1994.
- [3] J. Lu, M. Huang, J. Yang, and P. Li, *Wireless Personal Communications*, Springer Netherlands, Dordrecht, Netherlands, 2020.

- [4] C. Wu, A. B. Khanikaev, R. Adato, N. Arju, A. Altug, and G. Shvets, "Fano-resonant asymmetric metamaterials for ultrasensitive spectroscopy and identification of molecular monolayers," *Nature Materials*, vol. 11, no. 1, p. 69, 2011.
- [5] E. A. Othman, A. N. Yusoff, M. Mohamad, H. Abdul Manan, A. I. Abd Hamid, and V. Giampietro, "Hemispheric lateralization of auditory working memory regions during stochastic resonance: an fMRI study," *Journal of Magnetic Resonance Imaging*, vol. 51, no. 6, p. 1821, 2020.
- [6] D. C. Mei, Z. L. Jia, and C. J. Wang, "Combined effects of asymmetry and noise correlation on the noise-enhanced stability phenomenon in a bistable system," *Physica Scripta*, vol. 84, no. 4, Article ID 045012, 2011.
- [7] F. Long, "Time-delay induced symmetry restoration and noise enhanced stability phenomena under correlated noises in an asymmetric bistable system," *Indian Journal of Physics*, vol. 88, no. 10, p. 1111, 2014.
- [8] Z.-L. Jia and D.-C. Mei, "The critical effects of time delay and noise correlation on stochastic resonance in an asymmetric bistable system," *Modern Physics Letters B*, vol. 25, no. 16, p. 1377, 2011.
- [9] C. Li, F. Long, and D. Mei, "Correlated effects of noise on symmetry of an asymmetric bistable system," *Frontiers in Physics*, vol. 10, no. 2, 2015.
- [10] F. Long, L. Du, and D. Mei, "Asymmetric effects on the associated relaxation time and the correlation function of a bistable system with correlated noises," *Physica Scripta*, vol. 79, no. 4, Article ID 045007, 2009.
- [11] J. J. Collins, C. C. Chow, and T. T. Imhoff, "Stochastic resonance without tuning," *Nature*, vol. 376, no. 6537, p. 236, 1995.
- [12] L. Gammaitoni, F. Marchesoni, and S. Santucci, "Stochastic resonance as a bona fide resonance," *Physical Review Letters*, vol. 74, no. 7, p. 1052, 1995.
- [13] M. I. Dykman and P. V. E. McClintock, "What can stochastic resonance do?" *Nature*, vol. 391, no. 6665, p. 344, 1998.
- [14] F. Moss and K. Wiesenfeld, "Stochastic resonance and the benefits of noise: from ice ages to crayfish and SQUIDS," *Nature*, vol. 373, pp. 33–36, 1995.
- [15] R. Benzi, A. Sutera, and A. Vulpiani, "The mechanism of stochastic resonance," *Journal of Physics A: Mathematical and General*, vol. 14, no. 11, pp. L453–L457, 1981.
- [16] B. Xu, F. Duan, R. Bao, and J. Li, "Stochastic resonance with tuning system parameters: the application of bistable systems in signal processing," *Chaos, Solitons & Fractals*, vol. 13, no. 4, pp. 633–644, 2002.
- [17] G. Hu, G. Nicolis, and C. Nicolis, "Periodically forced Fokker-Planck equation and stochastic resonance," *Physical Review A*, vol. 42, no. 4, pp. 2030–2041, 1990.
- [18] M. Evstigneev, P. Reimann, V. Pankov, and R. H. Prince, "Stochastic resonance in monostable overdamped systems," *Europhysics Letters (EPL)*, vol. 65, no. 1, p. 7, 2004.
- [19] J. M. G. Vilar and J. M. Rubí, "Divergent signal-to-noise ratio and stochastic resonance in monostable systems," *Physical Review Letters*, vol. 77, no. 14, p. 2863, 1996.
- [20] A. N. Grigorenko, S. I. Nikitin, and G. V. Roschepkin, "Stochastic resonance at higher harmonics in monostable systems," *Physical Review E*, vol. 56, no. 5, pp. R4907–R4910, 1997.
- [21] S. Lu, Q. He, H. Zhang, S. Zhang, and F. Kong, "Note: Signal amplification and filtering with a tristable stochastic resonance cantilever," *Review of Scientific Instruments*, vol. 84, no. 2, Article ID 026110, 2013.
- [22] P. Shi, X. Su, and D. Han, "Stochastic resonance in tristable system induced by dichotomous noise," *Modern Physics Letters B*, vol. 30, no. 31, p. 1650377, 2016.
- [23] S. Lu, Z. Dai, Y. Liu, G. Liu, H. Yang, and F. Wang, "Decreasing bit error rate by using enhanced tristable logical stochastic resonance in heavy noise condition," *Chinese Journal of Physics*, vol. 58, pp. 179–189, 2019.
- [24] L. Valerio, "Stochastic resonance for nonequilibrium systems," *Physical Review E*, vol. 100, p. 62124, 2019.
- [25] F. Guo, Y. R. Zhou, and Y. Zhang, "Stochastic resonance in a time-delayed bistable system subjected to dichotomous noise and white noise," *Chinese Journal of Physics*, vol. 48, no. 2, pp. 294–303, 2010.
- [26] A. M. D. Paor and B. Cogan, "Windspeed-dependent underdamping and its cure in the self-excited series-wound aerogenerator," *Applied Energy*, vol. 42, no. 4, pp. 253–267, 1992.
- [27] T. Kawamura, Y. Imamura, H. Motoshiromizu, and M. Abe, "Evaluation of overdamping behavior of a vibrating conductive cylinder in a strong magnetic field," *IEEE Transactions on Magnetics*, vol. 52, no. 10, p. 1, 2016.
- [28] N. V. Agudov, A. V. Krichigin, D. Valenti, and B. Spagnolo, "Stochastic resonance in a trapping overdamped monostable system," *Physical Review E*, vol. 81, no. 5 Pt 1, Article ID 051123, 2010.
- [29] L. Liu, L. Cao, and L. Zhang, "Stochastic resonance in an overdamped harmonic oscillator," *Acta Physica Sinica*, vol. 59, no. 3, pp. 1494–1498, 2010.
- [30] P. Landa, V. Ushakov, and J. Kurths, "Rigorous theory of stochastic resonance in overdamped bistable oscillators for weak signals," *Chaos, Solitons & Fractals*, vol. 30, no. 3, pp. 574–578, 2006.
- [31] P. Ushakov, S. C. Zhong, T. Zhe, and M. Hong, "Stochastic resonance of over-damped bistable system driven by chirp signal and Gaussian white noise," *Acta Physica Sinica*, vol. 62, no. 8, Article ID 080501, 2013.
- [32] B. Zhou and W. Xu, "Stochastic resonance in an asymmetric bistable system driven by multiplicative colored noise and additive white noise," *Chaos, Solitons & Fractals*, vol. 38, no. 4, p. 1146, 2008.
- [33] A. V. Barzykin and K. Seki, "Stochastic resonance driven by gaussian multiplicative noise," *Europhysics Letters (EPL)*, vol. 40, no. 2, p. 117, 1997.
- [34] P. H. Nggi, P. Jung, C. Zerbe, and F. Moss, "Can colored noise improve stochastic resonance?" *Journal of Statistical Physics*, vol. 70, p. 25, 1993.
- [35] B. Xu, J. Li, F. Duan, and J. Zheng, "Effects of colored noise on multi-frequency signal processing via stochastic resonance with tuning system parameters," *Chaos, Solitons & Fractals*, vol. 16, no. 1, p. 93, 2003.
- [36] X. Gu, "Stochastic resonance driven by time-delayed feedback in a bistable system with colored noise," *European Physical Journal D*, vol. 66, p. 67, 2012.
- [37] V. Sorokin and I. Blekhman, "On the stochastic resonance phenomenon in parametrically excited systems," *European Journal of Applied Mathematics*, vol. 30, no. 5, pp. 986–1003, 2018.
- [38] K. K. Wang, H. Ye, Y. J. Wang, and S. H. Li, "Impact of time delay and a multiplicative periodic signal on stochastic resonance and steady states shift for a stochastic insect outbreak system subjected to Gaussian noises," *Chinese Journal of Physics*, vol. 56, pp. 2204–2217, 2018.
- [39] G. Zhang, Y. J. Zhang, T. Q. Zhang, and R. Mdsheh, "Stochastic resonance in an asymmetric bistable system driven by

- multiplicative and additive Gaussian noise and its application in bearing fault detection,” *Chinese Journal of Physics*, vol. 56, pp. 1173–1186, 2018.
- [40] P. Shi, W. Zhang, D. Yuan, H. Xia, D. Han, and R. Fu, “Stochastic resonance in a time polo-delayed asymmetry bistable system driven by multiplicative white noise and additive color noise,” *Chaos Solitons & Fractals*, vol. 108, pp. 8–14, 2018.
- [41] H. Zhang, T. Yang, Y. Xu, and W. Xu, “Parameter dependence of stochastic resonance in the FitzHugh-Nagumo neuron model driven by trichotomous noise,” *European Physical Journal B*, vol. 88, p. 125, 2015.
- [42] B. Zhou and D. Lin, “Stochastic resonance in a time-delayed bistable system driven by trichotomous noise,” *Indian Journal of Physics*, vol. 91, no. 3, p. 299, 2017.
- [43] Q. Qi and B. Zhou, “Stochastic resonance in an underdamped periodic potential system with symmetric trichotomous noise,” *Indian Journal of Physics*, vol. 94, no. 1, pp. 81–86, 2019.
- [44] L. Lin, H. Wang, X. Huang, and Y. Wen, “Generalized stochastic resonance for a fractional harmonic oscillator with bias-signal-modulated trichotomous noise,” *International Journal of Modern Physics B*, vol. 32, p. 23, Article ID 1850072, 2017.
- [45] W. Zhao, L. Wang, and J. Fan, “Theory and method for weak signal detection in engineering practice based on stochastic resonance,” *International Journal of Modern Physics B*, vol. 31, no. 28, Article ID 1750212, 2017.
- [46] P.-M. Shi, Q. Li, and D.-Y. Han, “Stochastic resonance and MFPT in an asymmetric bistable system driven by correlated multiplicative colored noise and additive white noise,” *International Journal of Modern Physics B*, vol. 31, no. 14, Article ID 1750113, 2017.
- [47] A. A. Saha and G. V. Anand, “Design of detectors based on stochastic resonance,” *Signal Processing*, vol. 83, no. 6, pp. 1193–1212, 2003.

## Research Article

# Prediction Mathematic Model and Influencing Factors of Contact Stress of Cylindrical Gear with Arc Tooth

Qi Zhang <sup>1,2</sup>, Guang Wen,<sup>2</sup> Shuang Liang,<sup>3</sup> Qin Tong,<sup>4</sup> Li Hou,<sup>3</sup> and Guangchun Yang<sup>1</sup>

<sup>1</sup>School of Intelligent Manufacturing, Panzhihua College, Panzhihua 617000, China

<sup>2</sup>Equipment Manufacturing College, Chengdu Industrial Vocational Technical College, Chengdu 610218, China

<sup>3</sup>School of Mechanical Engineering, Sichuan University, Chengdu 610065, China

<sup>4</sup>School of Mechanical and Electrical Engineering, Mianyang Normal University, Mianyang 621000, China

Correspondence should be addressed to Qi Zhang; pzhuzq@stu.scu.edu.cn

Received 26 August 2020; Revised 3 October 2020; Accepted 12 October 2020; Published 3 November 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Qi Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Given the absence of a theoretical formula to analyze the influence of parameters on the contact stress of cylindrical gear with arc tooth, an explicit mathematical model of cylindrical gear with arc tooth between the design parameters and the contact stress is established based on Kriging surrogate model. The parameters of the variation function of Kriging model are optimized by using the whale optimization algorithm (WOA), and the explicit mathematical model accuracy between the design parameters and the contact stress of the gear is in turn optimized by the improved Kriging surrogate model. The influence of design parameters on the contact stress of cylindrical gear with arc tooth is analyzed based on the established mathematical model. The proposed algorithm was realized via the programming platform MATLAB; the simulation results indicate that the precision evaluation indexes (the correlation coefficient ( $R^2$ ), root mean square error (RMSE), and the relative maximum absolute error (RMAE)) of the proposed Kriging model are improved, in addition to the error range which is narrowed from  $(-2, 4)$  to  $(0, 3)$ . As the tooth width, modulus, pressure angle, and tooth line radius increased, the contact stress of the cylindrical gear with arc tooth gear declined, which was negatively correlated with the design parameter. The amplitude of contact stress of the cylindrical gear with arc tooth was the largest due to the change of tooth radius, followed by the change of modulus, while the influence of tooth width was less. Finally, the influence of modulus-tooth line radius interaction and pressure angle-tooth line radius interaction on contact stress of cylindrical gear with arc tooth was significant.

## 1. Introduction

Gears are widely used in various industries [1], and the involute cylindrical gear is widely used at present. However, in the involute cylindrical gear, the bearing capacity of spur gear is limited, the axial force of helical gear subsists, and the processing technology of herringbone gear is complex. In order to solve the problem of involute gear, Japanese scholar Yoshiro Hasegawa and others put forward a new type of circular arc gear transmission device. This new type of gear transmission has the advantages of good meshing performance, high coincidence, no axial force, stable transmission, and so on [2]. For this kind of gear, scholars have carried out in-depth research. Tseng et al. have studied its mathematical model, undercutting conditions, contact, and other aspect

[3–6]. Lei et al. have studied the accurate three-dimensional modeling method of circular tooth profile gear [7]. Dengqiu et al. have studied its contact distribution [8]. The contact performance of gears was studied by Chen Shuai and Wang [9, 10, 11, 12]. Surrogate model is an approximate modeling method that can generate mathematical models with a small number of samples to reduce the number of physical experiments and improve efficiency, and the commonly used surrogate model mainly includes response surface, Kriging, Support Vector Machine, Artificial Neural Network, etc. [13, 14]. In order to study the influence of design parameters on the contact stress of circular arc tooth line gear, this paper proposes using Kriging surrogate model to establish the mathematical model of contact stress between design parameters (tooth width, modulus, pressure angle, and tooth

radius) of cylindrical gear with arc tooth. The Kriging model with the best effect is selected by comparing the prediction effect of these common surrogate model methods on contact stress of cylindrical gear with arc tooth. With a view to improving the modeling accuracy of Kriging model, intelligent algorithm is used to optimize the variation function of Kriging; the commonly used intelligent algorithm mainly includes genetic algorithm, immune algorithm, simulated annealing algorithm, particle swarm algorithm, fish swarm algorithm, ant colony algorithm, cuckoo algorithm, and so forth [15]. In this paper, whale optimization algorithm (WOA) was used for the parameters of the variation function of the traditional Kriging model. The Kriging surrogate model with optimized parameters of the variation function is used to establish the explicit mathematical model between the design parameters and the contact stress of the gear. The influence of design parameters (tooth width, modulus, pressure angle, and tooth radius) on the contact stress of circular arc tooth line cylindrical gear is discussed based on the established mathematical model.

## 2. Mathematical Model of Circular Arc Gear

According to its forming principle [16, 17], the coordinate system of forming principle of circular arc tooth line cylindrical gear is shown in Figure 1. In Figure 1,  $S(O - XYZ)$  is the static coordinate system,  $S_1(O_1 - X_1Y_1Z_1)$  is the solidification coordinate of the gear blank, and  $S_T(O_T - X_TY_TZ_T)$  is the tool coordinate, which moves in relation to  $S(O - XYZ)$  coordinate at the speed of  $V_T = R \times \omega$ .

In the coordinate system  $S_T(O_T - X_TY_TZ_T)$ , the vector expression of the parametric equation of the surface formed in the cutting process is

$$\begin{aligned} \vec{r}_T = & -\left(\frac{\pm q \sin \alpha + R_T \pm \pi m}{4}\right) \cos \theta \vec{i}_T \\ & + \left(\frac{\pm q \sin \alpha + R_T \pm \pi m}{4}\right) \sin \theta \vec{j}_T + q \cos \alpha \vec{k}_T, \end{aligned} \quad (1)$$

$$\begin{cases} \vec{r}_{dq} = \pm \sin \alpha \cos \theta \vec{i}_T \pm \sin \alpha \sin \theta \vec{j}_T + \cos \alpha \vec{k}_T, \\ \vec{r}_{d\theta} = -\left(\frac{\pm q \sin \alpha + R_T + \pi m}{4}\right) \sin \theta \vec{i}_T + \left(\frac{\pm q \sin \alpha + R_T + \pi m}{4}\right) \cos \theta \vec{j}_T. \end{cases} \quad (4)$$

Then, the unit normal vector of the tool surface is

$$\begin{aligned} \vec{n} = \frac{\vec{r}_{dq} \times \vec{r}_{d\theta}}{|\vec{r}_{dq} \times \vec{r}_{d\theta}|} = & -\cos \alpha \cos \theta \vec{i}_T - \cos \alpha \sin \theta \vec{j}_T \\ & \pm \sin \alpha \vec{k}_T. \end{aligned} \quad (5)$$

where  $a$  is the pressure angle of the cutting edge of the cutting tool;  $\theta$  is the angle from the current enveloping point of the gear blank to the middle section; and  $q$  is the length of the enveloping reference point from the point on the cutting surface of the cutting tool along the direction of the generatrix of the cutting cone in the coordinate system.  $R$  is dividing circle radius of gear,  $R_T$  is cutter head radius of machining tool,  $R_n$  is inner edge radius in the tangent direction of dividing circle,  $R_w$  is outer edge radius in the tangent direction of dividing circle,  $R_w = R_T + \pi m/4$ ,  $m$  is gear module,  $\omega$  is cutter head rotation angle speed of machining tool, and  $V_T$  is cutter head moving speed of machining tool.

Transform  $S_T(O_T - X_TY_TZ_T)$  to  $S(O - XYZ)$ , and its coordinate conversion relationship is as follows:

$$\mathbf{T}_{OT} = \begin{bmatrix} 1 & 0 & 0 & R\varphi + R_T \\ 0 & 0 & 1 & R \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (2)$$

Transform  $S(O - XYZ)$  to  $S_1(O_1 - X_1Y_1Z_1)$ , and its coordinate conversion relationship is

$$\mathbf{T}_{10} = \begin{bmatrix} \cos \varphi & -\sin \varphi & 0 & 0 \\ \sin \varphi & \cos \varphi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (3)$$

If  $S_T(O_T - X_TY_TZ_T)$  is converted to  $S(O - XYZ)$ , then the coordinate transformation matrix is  $\mathbf{T}_{1T} = \mathbf{T}_{10} \mathbf{T}_{OT}$ .

**2.1. Unit Normal Vector of Tool Surface.** The tool surface equation is

**2.2. Relative Speed of Tool and Gear at Meshing Point.** The direction vector can be expressed as

$$\vec{\lambda} = \overline{O_1O_T} = (R\varphi + R_T) \vec{i}_T + R \vec{k}_T. \quad (6)$$

Therefore, the relative speed of the tool and the gear at the meshing point is



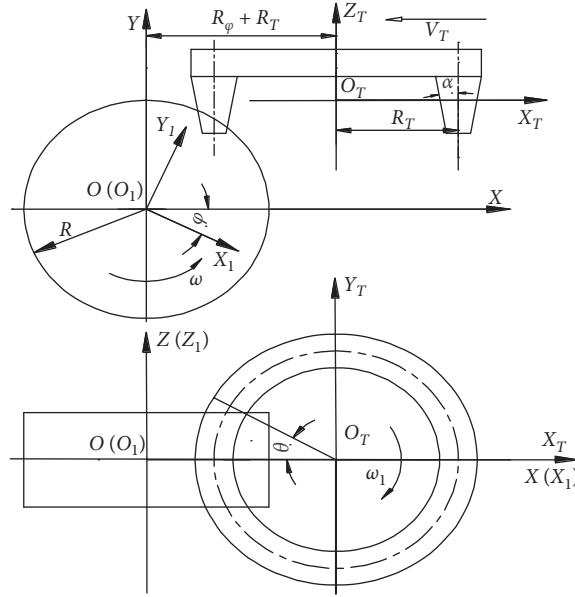


FIGURE 1: Coordinate system of forming principle of cylindrical gear with arc tooth.

$$\begin{aligned}
 \vec{v}^{12} &= \vec{\omega}^{12} \times \vec{r}^1 + \frac{d\lambda}{dt} - \vec{\omega}^2 \times \vec{\lambda}, \\
 &= -\omega_1 q \cos \alpha \vec{i}_T - \omega_1 \left( \frac{\pm q \sin \alpha + R_T + \pi m}{4} \right) \cos \theta \vec{k}_T + \omega_1 R \vec{i}_T \\
 &\quad - \left[ \omega_1 R \vec{i}_T - \omega_1 (R\phi + R_T) \vec{k}_T \right], \\
 &= -q \cos \alpha \omega_1 \vec{i}_T - \left[ \left( \frac{\pm q \sin \alpha + R_T + \pi m}{4} \right) \cos \theta + (R\phi + R_T) \right] \omega_1 \vec{k}_T.
 \end{aligned} \tag{7}$$

2.3. *Meshing Function.* Based on the meshing principle, the meshing function is expressed as follows:

$$\Gamma = \vec{n} \cdot \vec{v}^{12} = 0. \tag{8}$$

Because  $\omega_1 \neq 0$ , it can be concluded from the above formula that

$$q = \mp \sin \alpha \frac{\cos \theta (R_T + \pi m/4) + (R\phi + R_T)}{\cos \theta}. \tag{9}$$

2.4. *Conjugate Surfaces.* Combining equations (1) and (9), the contact line equation between the tool and the gear to be machined is as follows:

$$\begin{cases}
 x_T = -\left( \frac{\pm q \sin \alpha + R_T \pm \pi m}{4} \right) \cos \theta, \\
 y_T = \left( \frac{\pm q \sin \alpha + R_T \pm \pi m}{4} \right) \sin \theta, \\
 z_T = q \cos \alpha, \\
 q = \mp \sin \alpha \frac{\cos \theta (R_T + \pi m/4) + (R\phi + R_T)}{\cos \theta}.
 \end{cases} \tag{10}$$

By converting the coordinates in  $S_T(O_T - X_T Y_T Z_T)$  to  $S_1(O_1 - X_1 Y_1 Z_1)$ , the tooth surface equation of the cut gear can be obtained as follows:

$$\left\{ \begin{array}{l} x_1 = \left[ -\left( \frac{\pm q \sin \alpha + R_T \pm \pi m}{4} \right) \cos \theta + R\varphi + R_T \right] \cos \varphi - (q \cos \alpha + R) \sin \varphi, \\ y_1 = \left[ -\left( \frac{\pm q \sin \alpha + R_T \pm \pi m}{4} \right) \cos \theta + R\varphi + R_T \right] \sin \varphi + (q \cos \alpha + R) \cos \varphi, \\ z_1 = \left( \frac{\pm q \sin \alpha + R_T \pm \pi m}{4} \right) \sin \theta, \\ q = \mp \sin \alpha \frac{\cos \theta (R_T + \pi m/4) + (R\varphi + R_T)}{\cos \theta}. \end{array} \right. \quad (11)$$

2.5. *Instantaneous Contact Line.* Meshing function  $\Gamma = \Gamma(q, \theta, \varphi)$  is a function of variables  $q, \theta, \varphi$ . As  $\varphi_1 = \omega_1 t$ ,  $\varphi_1$  is regarded as a function of entering  $\Gamma$  in a certain instant, thus the function expression of the instantaneous contact line between the rack cutter and the gear blank can be obtained. If the value  $\varphi_1$  at any time is substituted into equation (7), the contact line equation of the whole surface can be obtained as follows:

$$\left\{ \begin{array}{l} x_T = -\left[ \left( 1 + \sin^2 \alpha \right) \left( \frac{R_T \pm \pi m}{4} \right) + \sin^2 \alpha \frac{R\varphi + R_T}{\cos \theta} \right] \cos \theta, \\ y_T = \left[ \left( 1 + \sin^2 \alpha \right) \left( \frac{R_T \pm \pi m}{4} \right) + \sin^2 \alpha \frac{R\varphi + R_T}{\cos \theta} \right] \sin \theta, \\ z_T = \mp \sin \alpha \cos \alpha \left( \left( \frac{R_T \pm \pi m}{4} \right) + \left( \frac{R\varphi + R_T}{\cos \theta} \right) \right). \end{array} \right. \quad (12)$$

2.6. *Tooth Profile Equation.* In the axial middle section of the gear, from its extended coordinate system, it can be seen that  $b=0$ ; then,  $\theta$  is also equal to zero. Substituting these two parameters into equation (11), the tooth profile equation of the middle section is obtained:

$$\left\{ \begin{array}{l} x_1 = \left[ -\left( \frac{\pm q \sin \alpha + R_T \pm \pi m}{4} \right) + R\varphi + R_T \right] \cos \varphi - (q \cos \alpha + R) \sin \varphi, \\ y_1 = \left[ -\left( \frac{\pm q \sin \alpha + R_T \pm \pi m}{4} \right) + R\varphi + R_T \right] \sin \varphi + (q \cos \alpha + R) \cos \varphi, \\ q = \mp \sin \alpha \left( \left( \frac{R_T + \pi m}{4} \right) + R\varphi + R_T \right). \end{array} \right. \quad (13)$$

According to equation (13), it can be seen that the tooth profile of the axial symmetry surface of the gear is involute.

Similarly, on the axial asymmetric plane, let  $z_1 = b$ ; according to the expressions of  $Z_1$  and  $q$  in equation (11), we can get

$$\varphi = \frac{\{ [(-b/\tan \theta) + (R_T \pm \pi m/4) \cos \theta] / \sin^2 \alpha - (R_T \pm \pi m/4) \cos \theta - R_T \}}{R} \quad (14)$$

Thus, the expression of the tooth profile of the non-intermediate section can be obtained as follows:

$$\left\{ \begin{array}{l} x_T = \left[ -\left( \frac{\pm q \sin \alpha + R_T \pm \pi m}{4} \right) \cos \theta + R\varphi + R_T \right] \cos \varphi - (q \cos \alpha + R) \sin \varphi, \\ y_T = \left[ -\left( \frac{\pm q \sin \alpha + R_T \pm \pi m}{4} \right) \cos \theta + R\varphi + R_T \right] \sin \varphi + (q \cos \alpha + R) \cos \varphi, \\ q = \mp \sin \alpha \frac{\cos \theta (R_T + \pi m/4) + (R\varphi + R_T)}{\cos \theta}, \\ \varphi = \frac{\{[-(b/\tan \theta) + (R_T \pm \pi m/4) \cos \theta] / \sin^2 \alpha - (R_T \pm \pi m/4) \cos \theta - R_T\}}{R}. \end{array} \right. \quad (15)$$

### 3. Contact Stress Analysis of Cylindrical Gear with Arc Tooth Based on Finite Element Method

**3.1. Model Material Attribute Definition.** Establish material information in ABAQUS, such as modulus of elasticity ( $E=2.08$  MPa), Poisson's ratio (0.298), and other parameters.

**3.2. Establishing Analytical Steps and Interactions.** Establish analysis steps, which mainly include defining analysis type (static), defining analysis increment step, determining iteration method, creating field variable, and historical variable and determining output parameters. Turn on nonlinearity, and define interaction as contact; then set contact type of finite element analysis of circular arc tooth line gear pair as "friction free." The MPC (multipoint constraint) is established by the rotating centers of the active and the driven.

**3.3. Applying Restraint and Moment.** The torque of magnitude  $6.08 \times 10^4$  N-mm is applied to the driving wheel. At the same time, MPC is added to the driving wheel: the direction of its rotation axis is set as free, and other rotation and translation are set as fixed. In the driven wheel, all directions of MPC constraints are fixed, and no motion is allowed in any direction.

**3.4. Gridding.** The method of sweeping is used to divide the mesh of gear pair, with C3D8I as the mesh type. The overall unit size of the gear is set to 2 mm, while the local subdivision size of the contact area is 0.02 mm. If the requirements are not met, continue to adjust the parameters. In the subsequent calculation, the mesh can be further refined to improve the quality of division. Through multiple trial calculation and comparative analysis results, the grid can be used as the final analysis grid if the calculation result is nearly stable. In the analysis process, due to the local subdivision of

the contact area, the number of grids for each pair of gears is about  $1.2 \times 10^6$ . The result of gear grid division is shown in Figure 2.

**3.5. Solution and Visualization.** Considering that the mesh number of each pair of gears is about  $1.2 \times 10^6$ , a parallel calculation is used in the solution. The calculated stress nephogram of driving wheel and that of driven wheel are shown in Figure 3. Figure 3(a) shows the dynamic contact stress diagram of driving wheel, and Figure 3(b) shows the contact stress diagram of driving wheel.

According to the simulation analysis, contact area is near dividing circle. As shown in Figure 3(a), the driving wheel of the contact area of the above standard pitch circle has a maximum stress value of 503.3 MPa. As shown in Figure 3(b), driven wheel contact area of the below standard pitch circle has a maximum stress value of 501 MPa. Therefore, the maximum contact stress is 503.3 MPa, and the difference in value between active and driven wheel is 2.3 MPa, only equal to 0.4570% of the maximum contact stress, which almost can be ignored. Similarly, it can be seen from Figures 3(a) and 3(b) that the contact area of the gear studied in this paper is an ellipse under the action of load, which confirms the fact that the contact area of point contact gear under the action of load is ellipse.

### 4. The Prediction Model of Contact Stress of Cylindrical Gear with Arc Tooth Based on Surrogate Model

**4.1. Test Design and Response Results.** The common test methods mainly include uniform test design, orthogonal test design, Latin square test design, and other methods. This paper studies the relationship between tooth width, modulus, pressure angle, tooth radius, and contact force. Four factors and three levels are selected for the simulation scheme by using the principle of regression orthogonal test method. The level table of gear design factors is shown in Table 1.

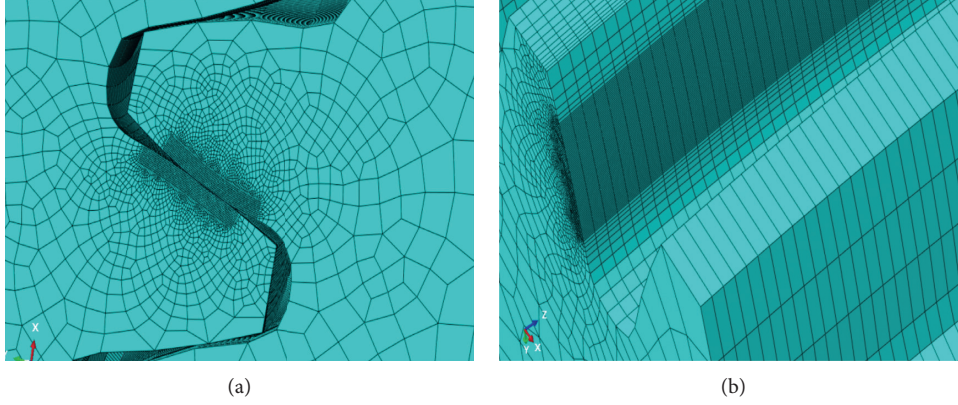


FIGURE 2: Meshing results of gears. (a) Global mesh results. (b) Local mesh results.

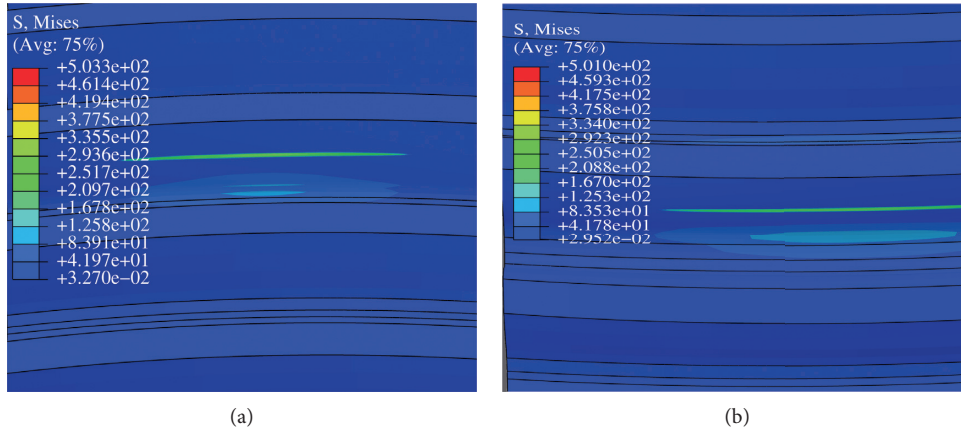


FIGURE 3: Stress neutrogram analysis of driving wheel and driven wheel. (a) Driving wheel. (b) Driven wheel.

TABLE 1: Gear design factor level.

| Level order | Tooth width (mm) | Modulus | Pressure angle (°) | Tooth radius (mm) |
|-------------|------------------|---------|--------------------|-------------------|
| Level 1     | 38.83685         | 2.96737 | 18.83685           | 191.8425          |
| Level 2     | 40               | 3.2     | 20                 | 250               |
| Level 3     | 41.16315         | 3.43263 | 21.16315           | 308.1575          |

In this paper, the orthogonal test is used to sample, and then the finite element method is used to get the gear joint stress under different sample data. According to the orthogonal table, 9 samples were processed, and the contact stress simulation results of  $L9$  (34) orthogonal test arc gear are shown in Table 2.

#### 4.2. Improved Kriging Model

**4.2.1. Kriging Surrogate Model.** Kriging model can obtain ideal fitting results when solving the problem of high nonlinearity. Its interpolation results are defined as the linear weighting of the response value of the known sample function, which is

$$y(x) = \sum_{j=1}^k \beta_j f_j(x) + Z(x). \quad (16)$$

In the formula,  $f_j(x)$  is function,  $\beta_j$  is corresponding coefficient, and  $Z(x)$  is static random process, whose mean value is 0, and its variance is  $\sigma^2$ . The covariance between the random variables corresponding to two different points in the design space is

$$\text{Cov}[Z(x^i), Z(x^j)] = \sigma^2 R(x^i, x^j), \quad (17)$$

$$R(x^i, x^j) = \prod_{l=1}^{N_s} R_l(\theta_l, x_l^i, x_l^j), \quad (18)$$

TABLE 2: Contact stress simulation results of L9 (34) orthogonal test.

| Order | Tooth width (mm) | Modulus | Pressure angle (°) | Tooth radius (mm) | Contact force (MPa) |
|-------|------------------|---------|--------------------|-------------------|---------------------|
| 1     | 38.83685         | 2.96737 | 18.83685           | 191.8425          | 542.7               |
| 2     | 38.83685         | 3.2     | 20                 | 250               | 457.2               |
| 3     | 38.83685         | 3.43263 | 21.16315           | 308.1575          | 396.8               |
| 4     | 40               | 2.96737 | 20                 | 308.1575          | 452.9               |
| 5     | 40               | 3.2     | 21.16315           | 191.8425          | 494.1               |
| 6     | 40               | 3.43263 | 18.83685           | 250               | 449.1               |
| 7     | 41.16315         | 2.96737 | 21.16315           | 250               | 469.6               |
| 8     | 41.16315         | 3.2     | 18.83685           | 308.1575          | 440.2               |
| 9     | 41.16315         | 3.43263 | 20                 | 191.8425          | 476.7               |

where  $R(x^i, x^j)$  is correlation function, which represents the correlation between random variables at different positions. The commonly used correlation function is Gaussian-type function.

In equation (17), the parameter  $\theta$  of the variation function of Kriging model is determined by the maximum likelihood estimation method to solve the optimization problem:

$$\max \left\{ -\frac{N_s}{2} \ln(\bar{\sigma}^2) - \frac{1}{2} \ln(R) - \frac{N_s}{2} \ln(2\pi) \right\}, \quad \theta > 0, \quad (19)$$

$$\bar{\sigma}^2 = \frac{1}{N_s} (y_s - F\tilde{\beta})^T R^{-1} (y_s - F\tilde{\beta}).$$

In order to minimize the root mean square error (RMSE) between Kriging prediction value and actual function value, the approximate expression of Kriging model can be obtained as

$$\tilde{y}(x) = f_x^T \tilde{\beta} + \left[ R(x, x^{(1)}), R(x, x^{(2)}), \dots, R(x, x^{(N_s)}) \right]^T R^{-1} (y_s - F\tilde{\beta}), \quad (20)$$

where  $F$  is matrix composed of basis function vector  $f_x^T$ ;  $\tilde{\beta}$  is least square estimate of KRG model coefficient;  $R$  is correlation matrix; and  $N_s$  is sample number.

**4.2.2. Whale Algorithm.** The principle of whale algorithm is derived from the "bubble net," foraging behavior of humpback whales. The mathematical expression based on this special strategy is as follows:

$$D = |C \cdot X^*(t) - X(t)|, \quad (21)$$

$$X(t+1) = X^*(t) - A \cdot D,$$

where  $t$  is the number of current iterations;  $X(t)$  is the coordinate vector of the current humpback whale;  $X(t+1)$  is the target coordinate vector after the next iteration;  $X^*(t)$  is the best position vector obtained so far, which is updated with time;  $D$  is the distance between the current humpback whale and the best position; and  $A$  and  $C$  are the number of systems, which are, respectively, expressed as

$$A = 2a \cdot r - a, \quad (22)$$

$$C = 2r,$$

where  $a$  is parameter in the value range (0, 2) and linearly decreasing with the iteration time and  $r$  is random vector in the interval (0, 1).

When  $|A| > 1$ , whales have the behavior of wandering and foraging. Using the random individual coordinates of the population to locate and navigate food, the mathematical expression is as follows:

$$X(t+1) = X_{\text{rand}}(t) - A \cdot D. \quad (23)$$

When  $|A| < 1$ , there were two behaviors of encircling and attacking prey, and the mathematical model is described as follows:

$$X(t+1) = X^*(t) - D \cdot e^{bl} \cdot \cos(2\pi l). \quad (24)$$

In the formula,  $b$  is the constant of the spiral shape and  $l$  is the random number on the interval  $(-1, 1)$ .

Since the contraction and spiral position updating of whales are carried out synchronously, mathematically selecting the same way of probability to update their position, then we can get the following expression:

$$X(t+1) = \begin{cases} X^*(t) - A \cdot D, & p < 0.5, \\ X^*(t) - D \cdot e^{bl} \cdot \cos(2\pi l), & p \geq 0.5. \end{cases} \quad (25)$$

According to the characteristics of Kriging and WOA algorithm, the Kriging surrogate model based on WOA algorithm is used for parameter impact analysis, and the flow chart is shown in Figure 4.

**4.3. Accuracy Evaluation Indexes of Kriging.** Generally, the evaluation indexes of correlation coefficient ( $R^2$ ), root mean square error (RMSE), and relative maximum absolute error (RMAE) are used to evaluate the accuracy of Kriging.

**4.3.1. The Correlation Coefficient ( $R^2$ ).**

$$R^2 = 1 - \frac{\sum_{i=1}^n (f_i - \hat{f}_i)^2}{\sum_{i=1}^n (f_i - \bar{f})^2}. \quad (26)$$

**4.3.2. The Root Mean Square Error (RMSE).**

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (f_i - \hat{f}_i)^2}{n}}. \quad (27)$$

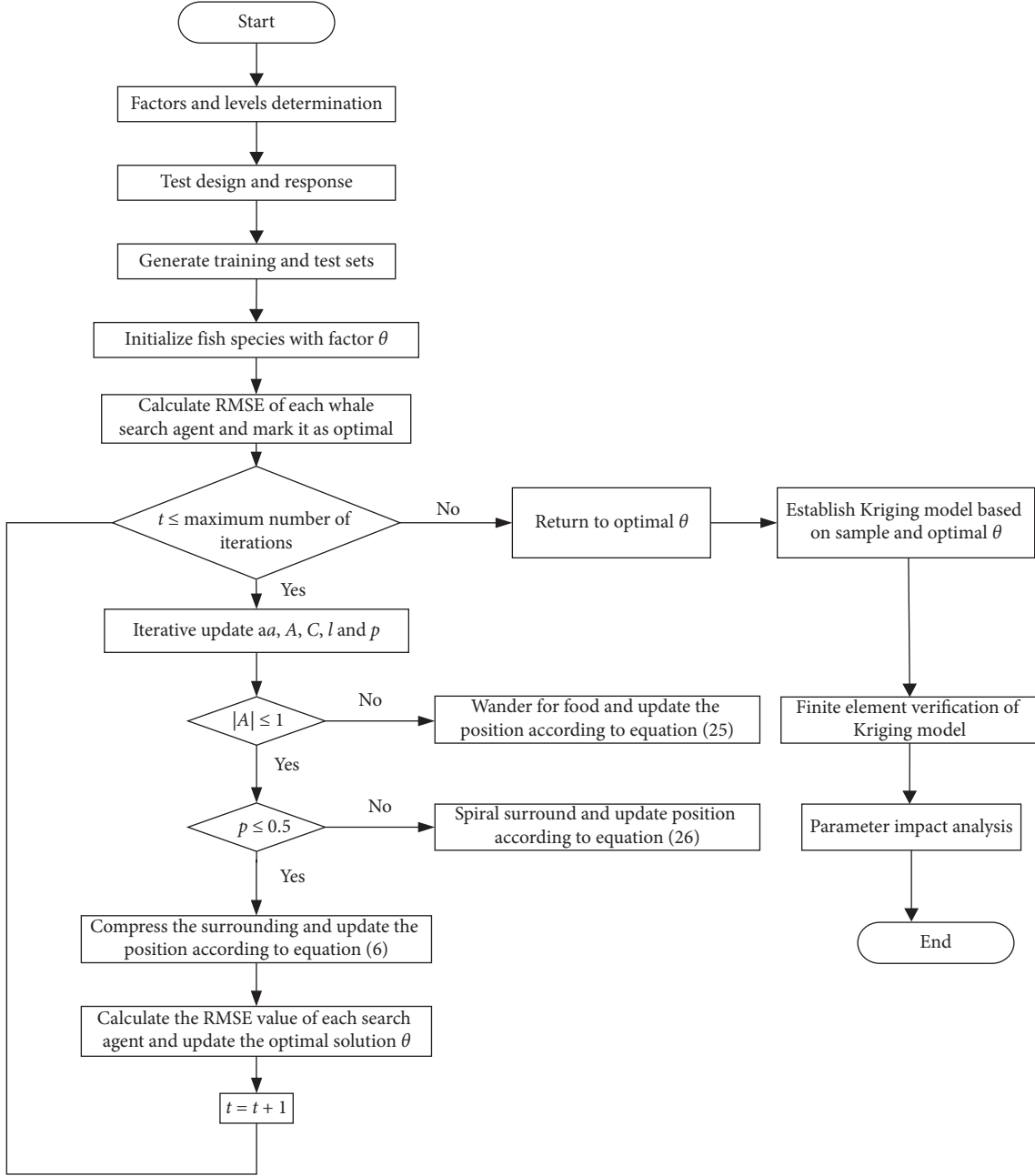


FIGURE 4: Improvement process of Kriging surrogate model based on WOA algorithm.

#### 4.3.3. The Relative Maximum Absolute Error (RMAE).

$$\text{RMAE} = \frac{\max\{|f_1 - \hat{f}_1|, |f_2 - \hat{f}_2|, \dots, |f_n - \hat{f}_n|\}}{\sqrt{1/n} \sum_{i=1}^n (f_i - \bar{f})^2}. \quad (28)$$

In equations (26) to (28),  $n$  is number of test sample points,  $f_i$  is the truth value of the actual model at the  $i^{\text{th}}$  test sample point, and  $\hat{f}_i$  is the predicted value of the surrogate model at the  $i^{\text{th}}$  test sample point. The closer  $R^2$  is to 1, the better the global approximation effect of the surrogate model will be. The smaller the RMSE is, the higher the precision of the surrogate

model will be, and the smaller the RMAE is, indicating that the higher the precision of the surrogate model will be.

4.4. Numerical Simulation. Using the MATLAB digital simulation platform, Kriging and the improved Kriging algorithm based on WOA algorithm are, respectively, used to establish the contact stress prediction model of circular arc gear. The evaluation indexes of the correlation coefficient ( $R^2$ ), root mean square error (RMSE), and the relative maximum absolute error (RMAE) before and after optimization are shown in Table 3.

TABLE 3: Evaluation indexes before and after optimization.

| Indexes | Kriging | WOA-Kriging | Optimization effect (%) |
|---------|---------|-------------|-------------------------|
| $R^2$   | 0.9922  | 0.9974      | 0.52                    |
| RMSE    | 2.8569  | 1.6540      | 42.11                   |
| RMAE    | 0.1322  | 0.0754      | 42.97                   |

It can be seen from Table 3 that the correlation coefficient ( $R^2$ ) has increased from 0.9922 to 0.9974, which is 0.52% higher and closer to 1. The correlation coefficient is closer to 1 implying that the global approximation ability of the optimized Kriging model is better. RMSE is used to indicate the impairment between the estimated value and the actual value. The higher the value of RMSE, the greater the impairment. After optimization, RMSE went from 2.8569 to 1.6540, reducing by 42.11%, which shows that the optimized Kriging model can better estimate the samples. After optimization, RMAE decreased from 0.1322 to 0.0754, which is 42.97% lower and closer to 0. This shows that the local error is less after optimization. Therefore, Kriging algorithm is improved based on WOA algorithm, which enhances the fitting ability and accuracy of Kriging algorithm.

The residual graph histogram of the test set for the contact stress prediction model is established by using Kriging and the improved Kriging algorithm based on WOA, as shown in Figures 5(a) and 5(b).

It can be seen from Figure 5 that the error dispersion of the optimized Kriging based on the improved WOA to establish the contact stress prediction model of the circular arc gear is more centralized, and the optimized error range is reduced from  $(-2, 4)$  to  $(0, 3)$ , which significantly improved accuracy.

**4.5. Finite Element Verification.** In order to further verify the validity of the established mathematical model, the finite element method is used to analyze the gear. The definition of material attribute of the model, the establishment of analysis steps and interactions, the application of constraints and moments, and mesh division are the same as those in Section 3. The gear parameters are shown in Table 4, and the results of the finite element analysis are shown in Figure 6.

As shown in Figure 6, the maximum contact stress of the gear is 472.2 MPa. The mathematical model established in Section 4.3 is used to predict the maximum contact stress of the gear in Table 4.

As shown in Table 5, the mathematical model established in Section 4.3 is used to predict the maximum contact stress of gears in Table 4. Compared with the results of finite element analysis, the maximum error is 3.2013 MPa, which shows that the mathematical model can effectively predict the contact stress of gears.

## 5. Influence of Design Parameters on Contact Stress

In this section, the influence of design parameters on contact stress is discussed based on the explicitly mathematical model between the gear design parameters (tooth width,

modulus, pressure angle, and tooth line radius) and the contact stress of cylindrical gear with arc tooth by using the proposed WOA-Kriging model. When the parameters (tooth width, modulus, pressure angle, and tooth line radius) are changed separately, the influence of tooth width, modulus, pressure angle, and tooth line radius on the contact stress of cylindrical gear with arc tooth is shown in Figure 7.

As seen in Figure 7, with the increase of tooth width, modulus, pressure angle, and tooth line radius, the contact stress of the cylindrical gear with arc tooth gear was in decline, which was negatively correlated with the design parameter, the amplitude of contact stress of the cylindrical gear with arc tooth is the largest due to the change of tooth radius, followed by the change of modulus, and the influence of tooth width is less.

The influence of change in tooth width, modulus, pressure angle, and tooth line radius on contact stress of the cylindrical gear with arc tooth is shown in Figure 8.

As seen in Figure 8, when tooth width interacts with modulus, pressure angle, and tooth line radius, respectively, the variation range of contact stress for the cylindrical gear with arc tooth gear is in decline. The response surface is not large and the complexity of surface is not high, indicating that the influence of tooth width interacting with modulus, pressure angle, and tooth line radius, respectively, on the contact stress is not significant. When modulus interacts with modulus, pressure angle, and tooth line radius, respectively, the complexity of the stress response surface is relatively high, which indicates that the influence on the stress is significant, compared with tooth width interacting with modulus, pressure angle, and tooth line radius, respectively. When pressure angle interacts with tooth line radius, the complexity of the stress response surface is relatively high, which also indicates that the influence on the stress is significant. Influence of modulus-tooth line radius interaction on contact stress of cylindrical gear with arc tooth and influence of pressure angle-tooth line radius interaction on contact stress of cylindrical gear with arc tooth are significant. In the contour map above, the smaller the distance between the contour lines is, the greater the influence of their interaction on the contact stress will be.

## 6. Conclusion

In order to study the influence of design parameters of cylindrical gear with arc tooth on its contact stress, a mathematical model of tooth width, modulus, pressure angle, radius of tooth line, and contact stress of gear is established by using Kriging surrogate model. Meanwhile,

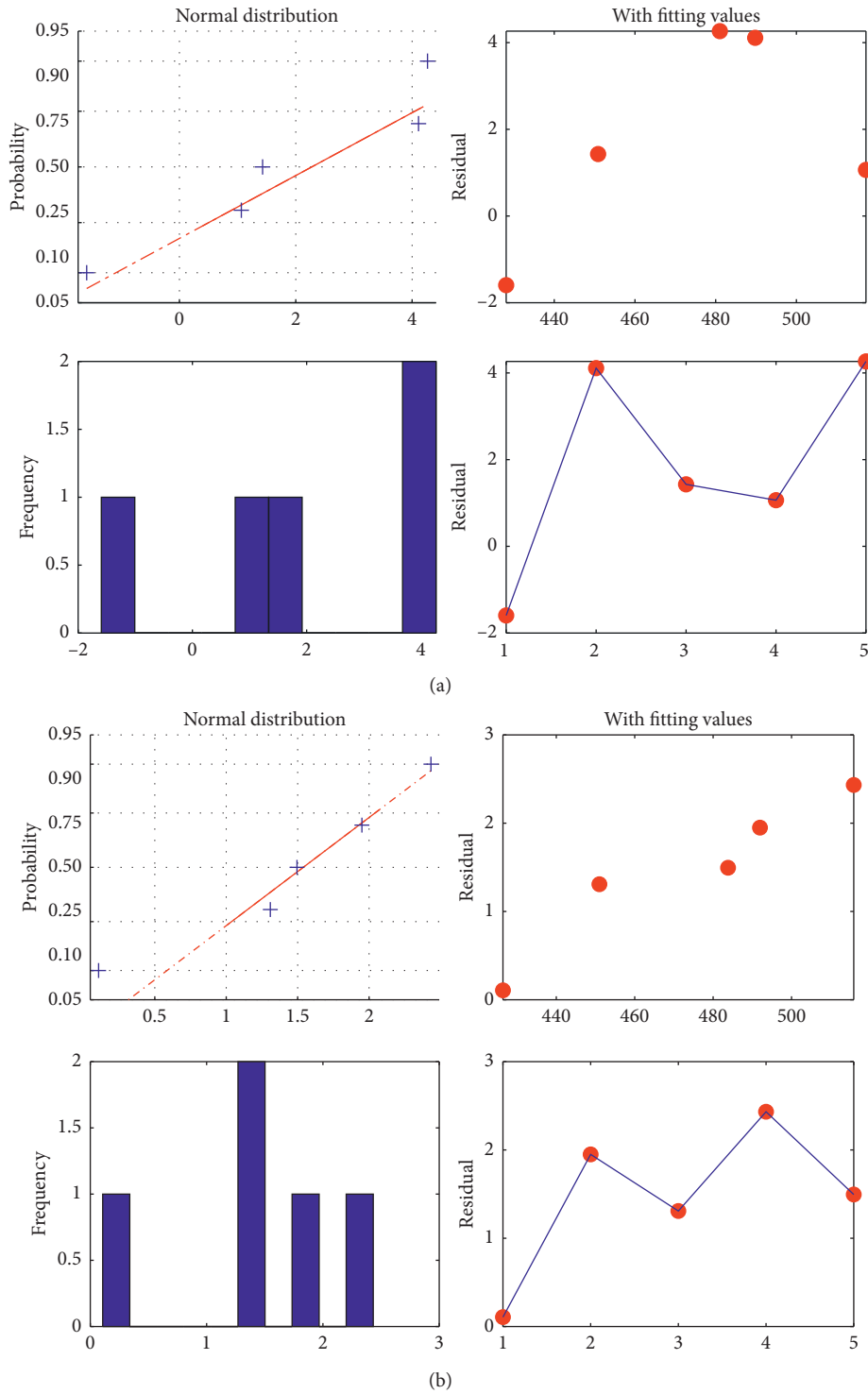


FIGURE 5: Residual diagram of contact stress prediction model. (a) Residual diagram of contact stress prediction model based on Kriging ((A) residual histogram, (B) residual with sequence, (C) residual, (D) sequence of observations). (b) Residual diagram of contact stress prediction model based on Kriging based on WOA ((A) residual histogram, (B) residual with sequence, (C) residual, (D) sequence of observations).

a Kriging model modeling method based on the whale optimization algorithm (WOA) is proposed. The parameters of variation function of traditional Kriging model are optimized by whale optimization calculation to improve the modeling accuracy of Kriging model. Based

on the established mathematical model, the influence of design parameters (tooth width, module, pressure angle, and tooth radius) of cylindrical gear with arc tooth on the contact stress of circular arc tooth line cylindrical gear is discussed.



TABLE 4: Gear parameters of finite element analysis.

| Order | Tooth width (mm) | Modulus  | Pressure angle (°) | Tooth radius (mm) |
|-------|------------------|----------|--------------------|-------------------|
| 1     | 38.836850        | 3.2      | 20                 | 308.157500        |
| 2     | 38.836850        | 2.967370 | 18.836850          | 308.157500        |
| 3     | 40               | 3.432630 | 21.163150          | 308.157500        |
| 4     | 40               | 3.432630 | 20.000000          | 191.842500        |
| 5     | 40               | 2.967370 | 21.163150          | 250               |

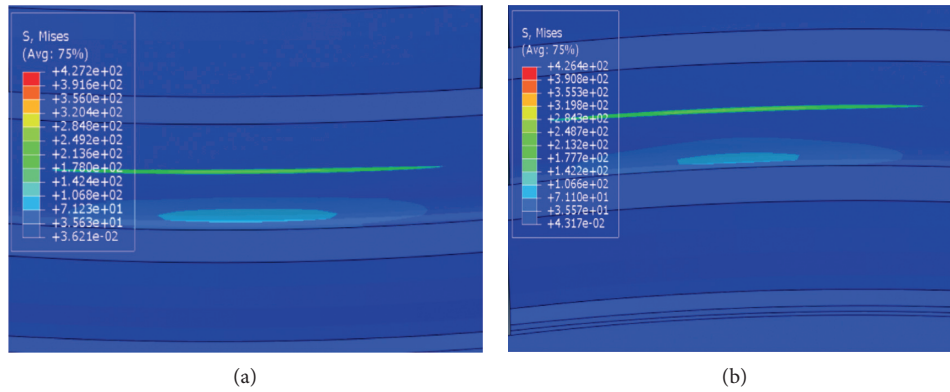


FIGURE 6: Stress nephogram of driving wheel and driven wheel. (a) Driving wheel. (b) Driven wheel.

TABLE 5: Finite element prediction results.

| Order | Tooth width (mm) | Modulus  | Pressure angle (°) | Tooth radius (mm) | Finite element stress (MPa) | Predicted stress (MPa) | Error (MPa) |
|-------|------------------|----------|--------------------|-------------------|-----------------------------|------------------------|-------------|
| 1     | 38.83685         | 3.2      | 20                 | 308.157500        | 427.2                       | 429.7935               | 2.5935      |
| 2     | 38.83685         | 2.967370 | 18.836850          | 308.157500        | 462.3                       | 465.5013               | 3.2013      |
| 3     | 40               | 3.432630 | 21.163150          | 308.157500        | 396.4                       | 394.7222               | -1.6778     |
| 4     | 40               | 3.432630 | 20.000000          | 191.842500        | 479.2                       | 478.3456               | -0.8544     |
| 5     | 40               | 2.967370 | 21.163150          | 250               | 470                         | 471.8039               | 1.8039      |

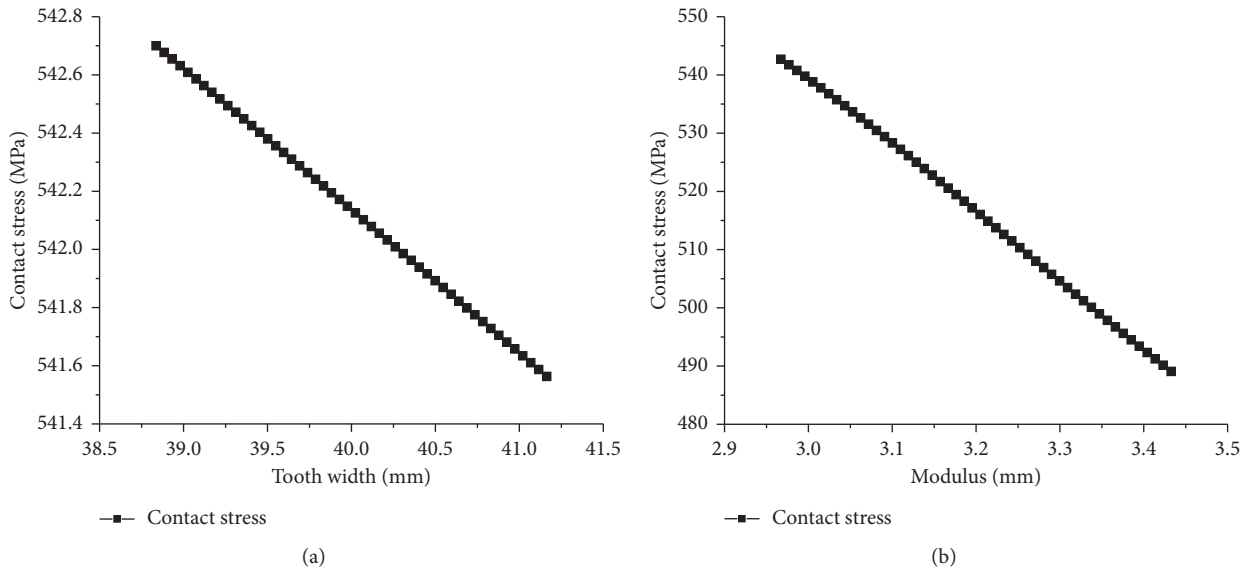


FIGURE 7: Continued.

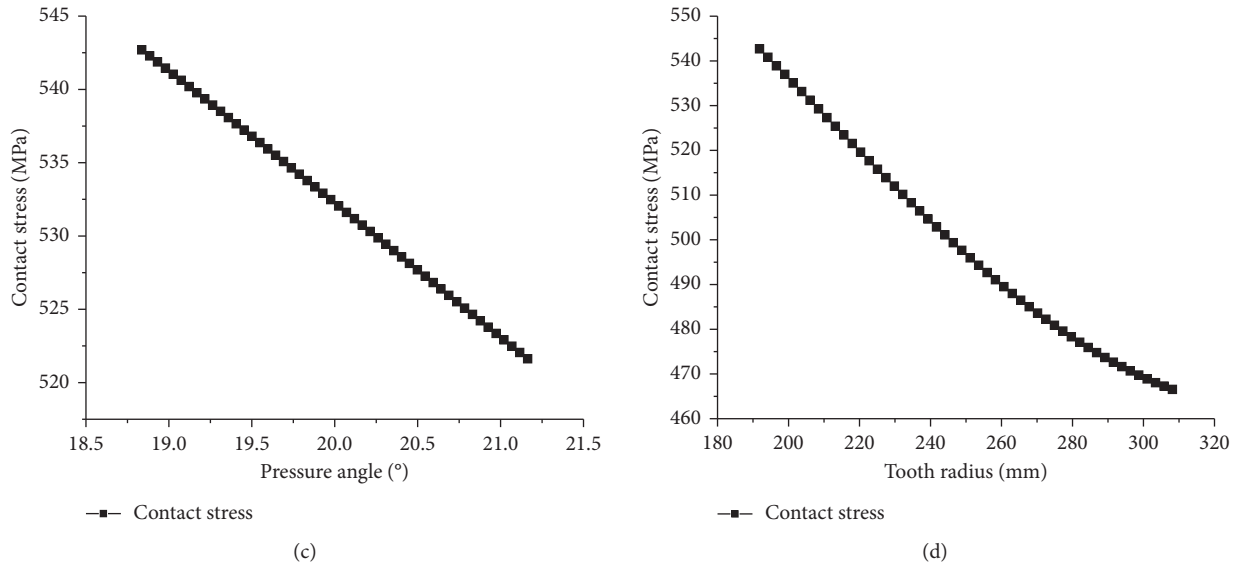


FIGURE 7: Influence of design parameters on contact stress. (a) Influence of tooth width on contact stress contact stress of cylindrical gear with arc tooth. (b) Influence of modulus on contact stress contact stress of cylindrical gear with arc tooth. (c) Influence of pressure angle on contact stress contact stress of cylindrical gear with arc tooth. (d) Influence of tooth line radius on contact stress contact stress of cylindrical gear with arc tooth.

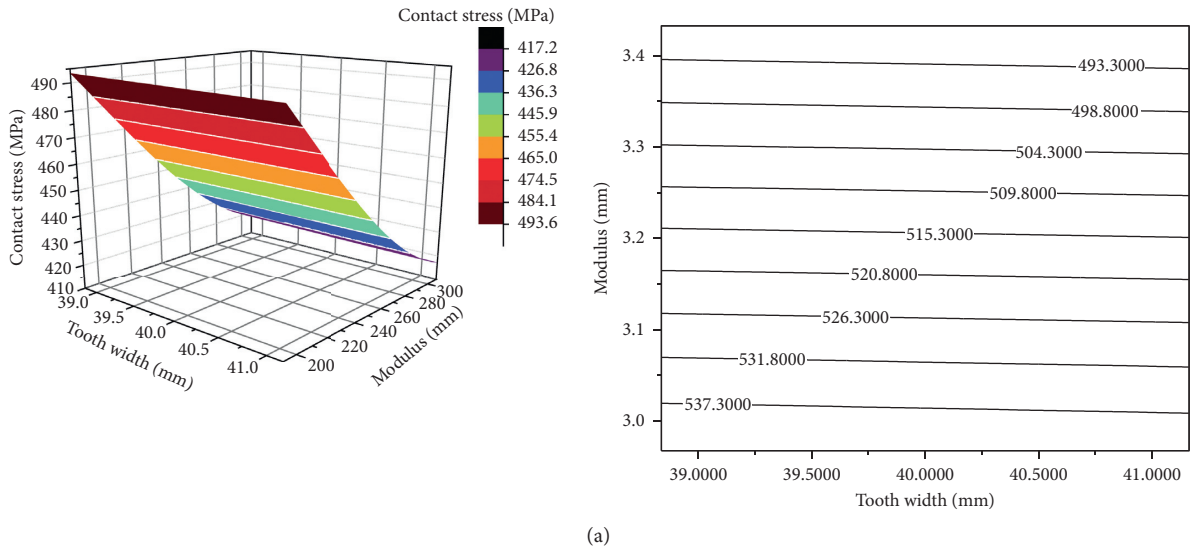
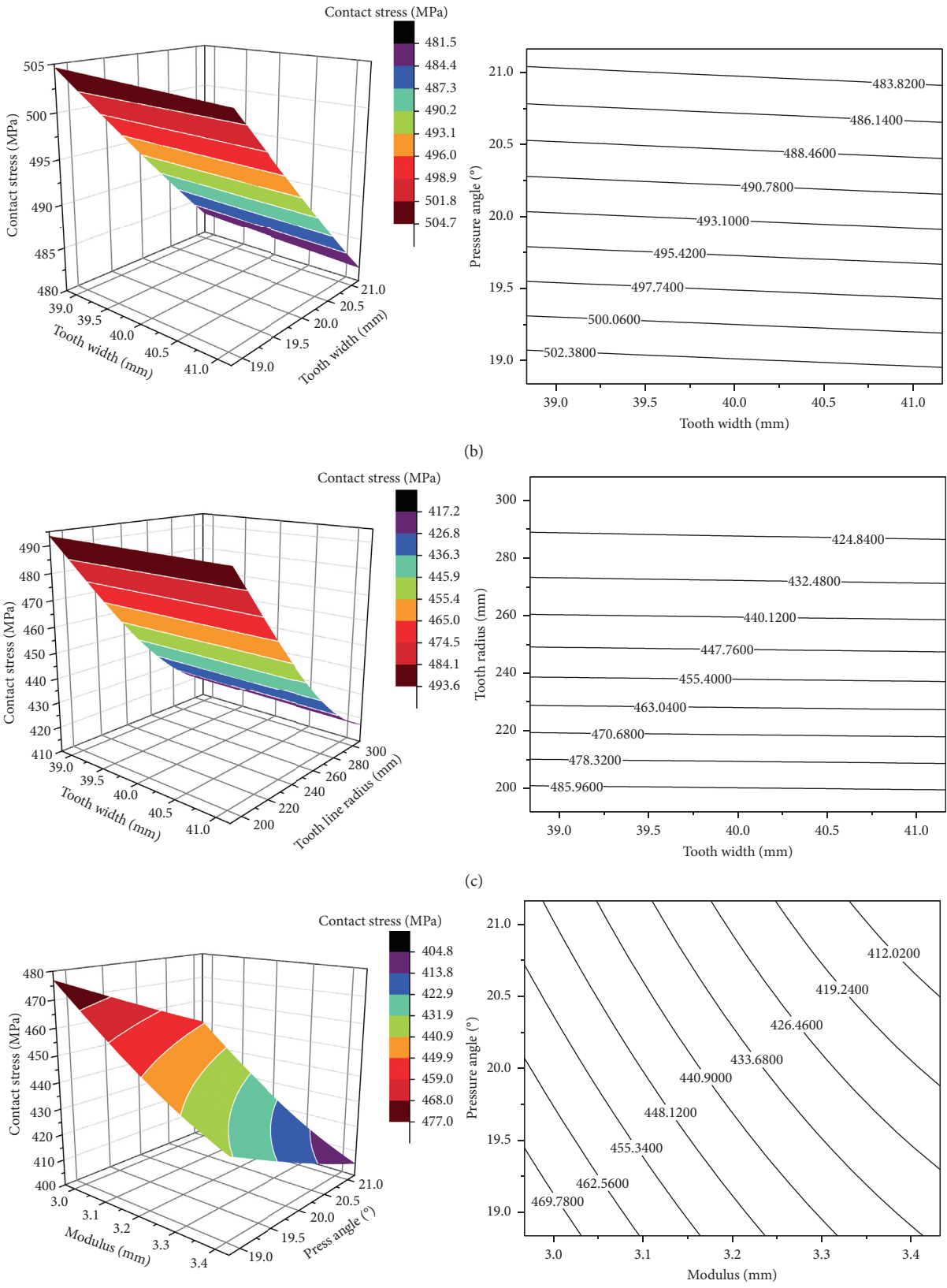


FIGURE 8: Continued.



(d)  
FIGURE 8: Continued.

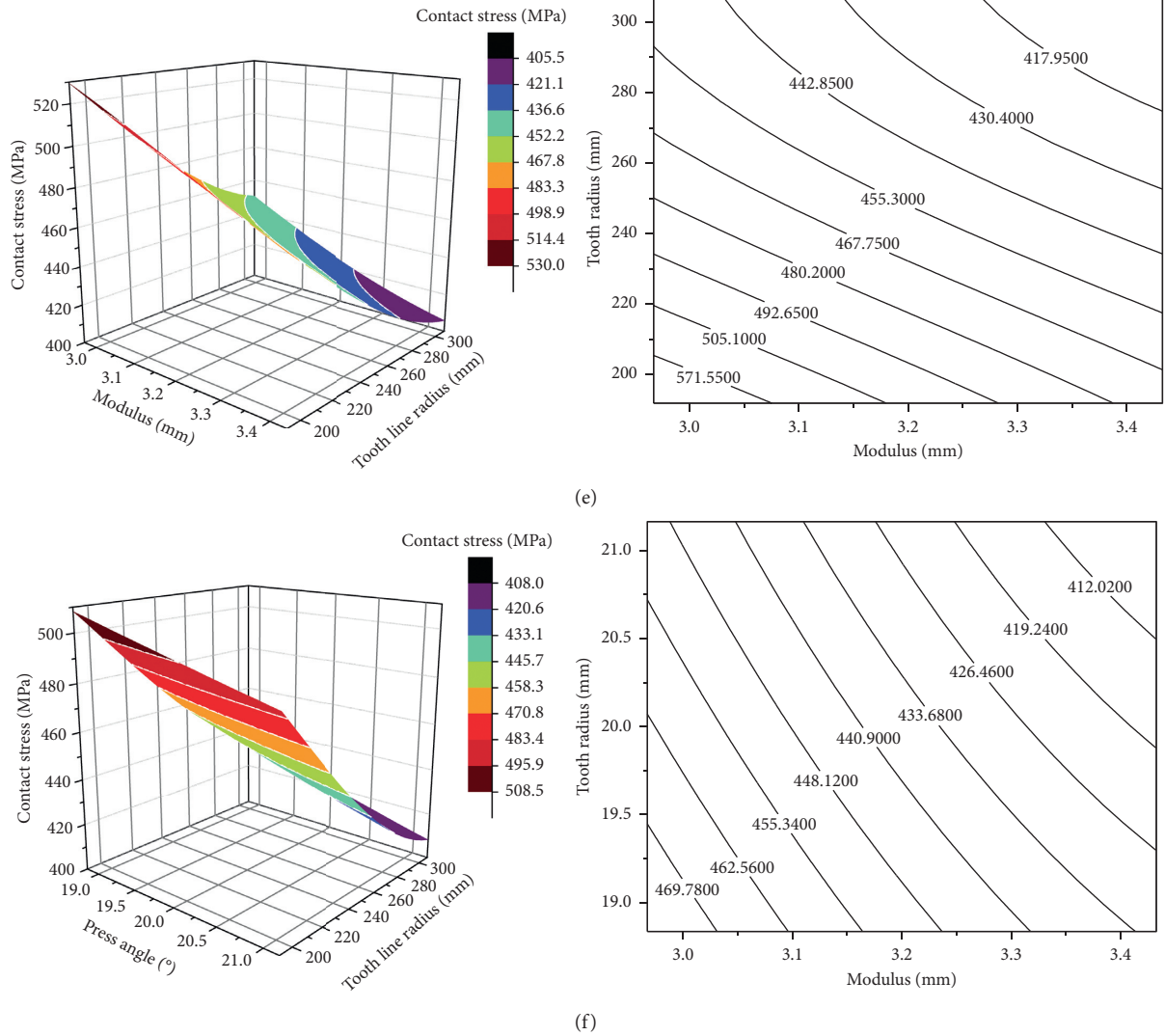


FIGURE 8: Influence of design parameters interaction on contact stress of cylindrical gear with arc tooth. (a) Influence of tooth width-modulus interaction on contact stress of cylindrical gear with arc tooth. (b) Influence of tooth width-pressure angle interaction on contact stress of cylindrical gear with arc tooth. (c) Influence of tooth width-tooth line radius interaction on contact stress of cylindrical gear with arc tooth. (d) Influence of modulus-pressure angle interaction on contact stress of cylindrical gear with arc tooth. (e) Influence of modulus-tooth line radius interaction on contact stress of cylindrical gear with arc tooth. (f) Influence of pressure angle-tooth line radius interaction on contact stress of cylindrical gear with arc tooth.

- (1) Based on the improved Kriging model of WOA algorithm, the correlation coefficient ( $R^2$ ) is increased from 0.9922 to 0.9974, MSE is reduced from 2.8569 to 1.6540, and RMAE is also reduced from 0.1322 to 0.0754. The correlation coefficient ( $R^2$ ) of the improved algorithm is improved, and root mean square error (RMSE) and relative maximum absolute error (RMAE) are improved to some extent, which improves the global approximation ability of the traditional Kriging algorithm, reduces the local error, and improves the fitting accuracy.
- (2) Both the traditional Kriging algorithm and the improved Kriging algorithm based on WOA algorithm can establish the contact stress prediction model of gear, and the accuracy is within the applicable range.

The error range of Kriging is  $(-2, 4)$ , while the error range of Kriging algorithm based on WOA algorithm is  $(0, 3)$ . The accuracy of the improved Kriging has been optimized obviously, and the accuracy of the model is verified by the finite element method, where the maximum error is only 2.5935 MPa.

- (3) Based on the established mathematical model, the influence of design parameters (tooth width, modulus, pressure angle, and tooth radius) on the contact stress of circular arc tooth line cylindrical gear is discussed. With the increase of tooth width, modulus, pressure angle, and tooth line radius, the contact stress of the cylindrical gear with arc tooth gear was in decline, which was negatively correlated with the design parameter, the amplitude of contact

stress of the cylindrical gear with arc tooth is the largest due to the change of tooth radius, followed by the change of modulus, and the influence of tooth width is the least. And influence of modulus-tooth line radius interaction and pressure angle-tooth line radius interaction on contact stress of cylindrical gear with arc tooth is significant.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no known conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This project was supported by National Natural Science Foundation of China (nos. 51875370 and 51375320), by Science and Technology Research Program of Sichuan Province (no.2018JY0420), by the seed fund of University Science Park of Panzhihua City (no. University Science Park 2019-25), and by University Key Laboratory of Sichuan in Process Equipment and Control Engineering (GK201905).

## References

- [1] J. Li, Q. J. Sun, and L. Q. Wang, "Tooth contact analysis of involute gear," *Chinese Journal of Engineering Design*, vol. 16, no. 1, pp. 27–31, 2009.
- [2] S. Hasegawa, "Cutting machine of cylindrical gears with curvilinear shaped teeth," *Precision Machine*, vol. 18, no. 8, pp. 253–257, 1952.
- [3] J. T. Tseng and C. B. Tsay, "Mathematical model and surface deviation of cylindrical gears with curvilinear shaped teeth cut by a hob cutter," *Journal of Mechanical Design*, vol. 127, no. 5, pp. 982–987, 2005.
- [4] R.-T. Tseng and C.-B. Tsay, "Contact characteristics of cylindrical gears with curvilinear shaped teeth," *Mechanism and Machine Theory*, vol. 39, no. 9, pp. 905–919, 2004.
- [5] J.-T. Tseng and C.-B. Tsay, "Undercutting and contact characteristics of cylindrical gears with curvilinear shaped teeth generated by hobbing," *Journal of Mechanical Design*, vol. 128, no. 3, pp. 634–643, 2006.
- [6] A. Fuentes, R. Ruiz-Orzaez, and I. Gonzalez-Perez, "Computerized design, simulation of meshing, and finite element analysis of two types of geometry of curvilinear cylindrical gears," *Computer Methods in Applied Mechanics and Engineering*, vol. 272, pp. 321–339, 2014.
- [7] C. Lei, H. Li, Z. Fei et al., "Accurate three-dimensional modeling of cylindrical gear with variable hyperbolic circular-arc-tooth-trace," *Machinery Design & Manufacture*, vol. S1, pp. 54–56, 2019.
- [8] D. Dengqiu, Y. Zhenhuan, F. Qianjun, and A. Yu, "Research on contact stress distribution and meshing position change rule of variable hyperbolic cylindrical gear with arcuate tooth trace," *Design and Research*, vol. 9, pp. 53–58, 2019.
- [9] C. Shuai, H. Li, T. Meng et al., "Dynamic contact characteristic analysis of cylindrical gear with variational hyperbola and circular-arc-tooth-trace," *Journal of Mechanical Transmission*, vol. 42, no. 9, pp. 85–89, 2018.
- [10] W. Hong, H. Li, Z. Fei, H. Lin-Tong, Z. Qi-Fan, and Q. Xin, "Dynamic contact analysis of cylindrical gear with arcuate tooth trace," *Modular Machine Tool & Automatic Manufacturing Technique*, vol. 6, pp. 37–40, 2017.
- [11] Z. Haiyan, H. Li, W. Yongqiao, H. Linli, and L. Bo, "Modeling and contact strength analysis for modified cylindrical gear profile with arcuate tooth trace," *Journal of Mechanical Transmission*, vol. 39, no. 7, pp. 20–22, 2015.
- [12] L. Bo, H. Li, J. Ping et al., "Modeling and contact strength analysis of cylindrical gear with arcuate tooth trace," *Journal of Mechanical Transmission*, vol. 38, no. 2, pp. 100–105, 2014.
- [13] A. K. Gupta, "Predictive modelling of turning operations using response surface methodology, artificial neural networks and support vector regression," *International Journal of Production Research*, vol. 48, no. 3-4, pp. 763–778, 2010.
- [14] B. Gaspar, A. P. Teixeira, and C. G. Soares, "Assessment of the efficiency of kriging surrogate models for structural reliability analysis," *Probabilistic Engineering Mechanics*, vol. 37, pp. 24–34, 2014.
- [15] V. I. Skoullis, I. X. Tassopoulos, and G. N. Beligiannis, "Solving the high school timetabling problem using a hybrid cat swarm optimization based algorithm," *Applied Soft Computing*, vol. 52, pp. 277–289, 2016.
- [16] L. Qin, L. Haidong, and Z. Qi, "Study on the tooth surface equation and parametric modeling for new type of cylindrical gear with arc tooth trace," *Journal of Mechanical Transmission*, vol. 40, no. 8, pp. 50–53, 2016.
- [17] D. Fayun and Z. Qi, "Mathematical model and modal analysis of new type of circular arc tooth trace cylindrical gear," *Journal of Mechanical Transmission*, vol. 42, no. 1, pp. 22–26, 2018.

## Research Article

# Analysis of Asymmetric Piecewise Linear Stochastic Resonance Signal Processing Model Based on Genetic Algorithm

Lina He <sup>1</sup> and Chuan Jiang <sup>2</sup>

<sup>1</sup>Chongqing College of Electronic Engineering, Chongqing 400031, China

<sup>2</sup>Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Correspondence should be addressed to Chuan Jiang; [s180101111@stu.cqupt.edu.cn](mailto:s180101111@stu.cqupt.edu.cn)

Received 23 August 2020; Revised 12 October 2020; Accepted 15 October 2020; Published 30 October 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Lina He and Chuan Jiang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The stochastic resonance system has the advantage of making the noise energy transfer to the signal energy. Because the existing stochastic resonance system model has the problem of poor performance, an asymmetric piecewise linear stochastic resonance system model is proposed, and the parameters of the model are optimized by a genetic algorithm. The signal-to-noise ratio formula of the model is derived and analyzed, and the theoretical basis for better performance of the model is given. The influence of the asymmetric coefficient on system performance is studied, which provides guidance for the selection of initial optimization range when a genetic algorithm is used. At the same time, the formula is verified and analyzed by numerical simulation, and the correctness of the formula is proved. Finally, the model is applied to bearing fault detection, and an adaptive genetic algorithm is used to optimize the parameters of the system. The results show that the model has an excellent detection effect, which proves that the model has great potential in fault detection.

## 1. Introduction

When the nonlinear system interacts with the signal to be measured and the nonsignal to be measured (noise), resonance will occur. At this time, when the noise intensity increases, the noise will not inundate the signal but will increase the signal-to-noise ratio, and the noise energy will transfer to the signal. This shows that stochastic resonance is a powerful method to extract a weak signal from strong noise [1, 2]. However, a large number of studies have shown that the classical SR theory has obvious detection advantages only in the case of small parameters, i.e., adiabatic approximation. In engineering application, most target signals are not small parameter signals, which greatly limits the application in engineering. In recent years, a series of achievements have been made in the resonance phenomenon of large parameter signals, such as the use of secondary sampling, single sideband modulation, frequency-domain information exchange, and other methods, which can make large parameter signals produce resonance phenomenon [3]. In reference

[4], it is known that the piecewise linear model has better performance than the classical model, and in reference [5], the asymmetric model can obtain better performance than the symmetric model by adjusting the asymmetric factor. Zhang et al. applied a stochastic resonance system for bearing fault detection [6, 7]. Zhang et al. found that the performance of the system is also affected by the system parameters [8], and the selection of parameters will directly affect the quality of the system. Zhang and He obtained better system parameters through an adaptive genetic algorithm and applied them to bearing fault detection [9].

At present, some scholars are committed to exploring the effect of asymmetry on system performance [2, 4]. Some scholars have devoted themselves to the study of piecewise linear systems. Both of them can achieve better system performance [10]. However, there is no research on the combination of them to obtain better performance. Based on this, an asymmetric linear piecewise bistable model is proposed. The analytical expression and signal-to-noise ratio of the model are derived and compared with the symmetrical

bistable piecewise linear stochastic resonance system and the continuous bistable system. At the same time, in order to obtain better performance in the application of bearing fault detection, a genetic algorithm is used to obtain better system parameters. In the second part of the paper, the concrete formula derivation is given. It provides a theoretical basis for the simulation and experiment. In the third part, numerical simulation is carried out to verify the correctness of the formula derivation. In the fourth part, the genetic algorithm is proposed to optimize the parameters, which can improve the system performance to the greatest extent. Finally, it is applied to fault detection. The results show that this method can effectively extract fault feature frequency and has better performance than the symmetric model.

## 2. Theoretical Analysis of Asymmetric Piecewise Linear Model

*2.1. Asymmetric Piecewise Linear Model.* The expression of potential function  $U(x)$  is as follows:

$$U(x) = \begin{cases} \frac{k_3}{r(k_1 - k_2)}(x + rk_1), & x < -rk_2, \\ \frac{k_3}{rk_2}x, & -rk_2 \leq x < 0, \\ \frac{k_3}{k_2}x, & 0 \leq x < k_2, \\ -\frac{k_3}{(k_1 - k_2)}(x - k_1) & k_2 \leq x. \end{cases} \quad (1)$$

where  $D$  is the noise intensity,  $r$  is asymmetric factor,  $k_1$ ,  $k_2$  and  $k_3$  are system parameters. The transition rate  $n_{\pm}(t)$  is generally considered to have an exponential form. Under the action of periodic signal  $s(t) = A \cos(\omega_0 t)$ , it is expanded by Taylor series, as shown in the following equation:

$$\begin{aligned} n_+ &= \tilde{\lambda}_+ \left[ 1 + \frac{Ax_m}{D} \cos(\omega_0 t) + \frac{1}{2} \left( \frac{Ax_m}{D} \right)^2 \cos^2(\omega_0 t) + \dots \right], \\ n_- &= \tilde{\lambda}_- \left[ 1 - \frac{Ax_n}{D} \cos(\omega_0 t) + \frac{1}{2} \left( \frac{Ax_n}{D} \right)^2 \cos^2(\omega_0 t) - \dots \right]. \end{aligned} \quad (3)$$

Under the assumption of adiabatic approximation, the probability equation of the model can be established according to (2) and (3):

Among them,  $r$  is asymmetric factor,  $k_1$ ,  $k_2$  and  $k_3$  are system parameters and are all greater than 0. Under the static condition, the system has two potential wells and one barrier. The bottom of the two wells is  $x_+ = k_2$  and  $x_- = -rk_2$ , respectively, and the barrier height is  $\Delta U = c$ .

When  $r = 1$  is a symmetric piecewise linear model, other cases are asymmetric linear systems. It can be seen from Figure 1 that the width of a potential well varies with the asymmetric factor.

*2.2. System Response and Signal-to-Noise Ratio.* The steady state of the system is  $+x_m$ ,  $-x_n$  and  $rx_m = x_n$ . Let  $W_{\pm}(t)$  be the probability of time  $t$  system in bistability  $+x_m$ ,  $-x_n$  and define  $n_{\pm}(t)$  as the probability of transition from steady state  $+x_m$ ,  $-x_n$  at time  $t$ . Because of the asymmetry of the two potential wells, the transition probability of the two potential wells is not the same, that is,  $n_+(t) \neq n_-(t)$ . When the asymmetry factor is  $r = 1$ , the equation holds. According to the adiabatic approximation theory [2], the escape rate of the asymmetric linear bistable system is as follows:

$$\tilde{\lambda}_-^{-1} = \frac{1}{D} \int_{-rk_1}^{-rk_2} e^{-(1/D)((k_3/r(k_1 - k_2))(x + rk_1))} dx \int_{-rk_2}^0 e^{(1/D)((k_3/rk_2)x)} dx, \tilde{\lambda}_+^{-1} = \frac{1}{D} \int_{-k_1}^{-k_2} e^{-(1/D)((k_3/(k_1 - k_2))(x + k_1))} dx \int_{-k_2}^0 e^{(1/D)((k_3/k_2)x)} dx, \quad (2)$$

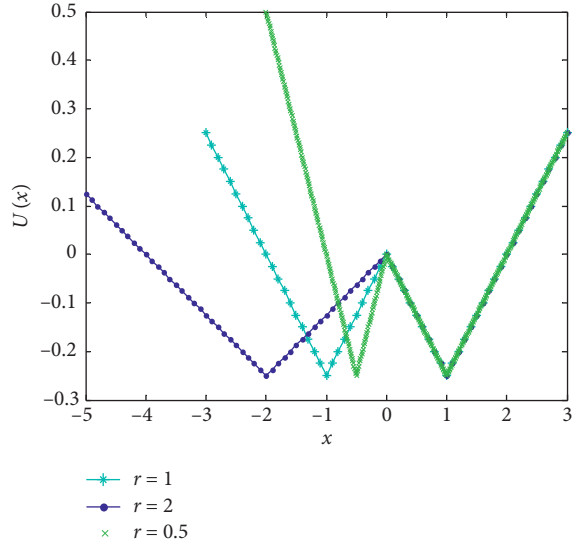
$$\frac{dw_-(t)}{dt} = -\tilde{\lambda}_- W_-(t) + \tilde{\lambda}_+ W_+(t), \quad (4)$$

$$\frac{dw_+(t)}{dt} = -\tilde{\lambda}_+ W_+(t) + \tilde{\lambda}_- W_-(t).$$

The equations of solution (4) can be obtained as follows:

$$\begin{aligned} W_-(t) &= \frac{1}{2} \left[ \frac{1}{\tilde{\lambda}_-} + e^{-(\tilde{\lambda}_- + \tilde{\lambda}_+)t} \right], \\ W_+(t) &= \frac{1}{2} \left[ \frac{1}{\tilde{\lambda}_+} - e^{-(\tilde{\lambda}_- + \tilde{\lambda}_+)t} \right]. \end{aligned} \quad (5)$$

The probability distribution function obtained from formula (5) is shown in the following equation:

FIGURE 1: Potential function  $U(x)$ .

$$\begin{aligned} \omega_-(t) &= \dot{W}_-(t) = -\left(\frac{\bar{\lambda}_- + \bar{\lambda}_+}{2}\right) e^{-(\bar{\lambda}_- + \bar{\lambda}_+)t}, \\ \omega_+(t) &= \dot{W}_+(t) = \left(\frac{\bar{\lambda}_- + \bar{\lambda}_+}{2}\right) e^{-(\bar{\lambda}_- + \bar{\lambda}_+)t}. \end{aligned} \quad (6)$$

The frequency-domain probability distribution function can be obtained by Fourier transform of formula (6)

$$\begin{aligned} |w(\omega)| &= |w_-(\omega)| + |w_+(\omega)| \\ &= \frac{\bar{\lambda}_- + \bar{\lambda}_+}{\sqrt{(\bar{\lambda}_- + \bar{\lambda}_+)^2 + \omega^2}}, \\ \varphi(\omega) &= \arctan\left(\frac{\omega}{\bar{\lambda}_- + \bar{\lambda}_+}\right). \end{aligned} \quad (7)$$

Since  $A$  is small, the response of the bistable system can be expressed as follows:

$$\langle x(t) \rangle = \bar{x}(D) \cos(\omega_0 t + \bar{\phi}(D)), \quad (8)$$

where

$$\bar{x}(D) = \frac{(1+r^2)Ax_m^2}{D} \frac{\bar{\lambda}_- + \bar{\lambda}_+}{\sqrt{(\bar{\lambda}_- + \bar{\lambda}_+)^2 + \omega_0^2}}, \quad (9)$$

$$\bar{\phi}(D) = \arctan\left(\frac{\omega_0}{\bar{\lambda}_- + \bar{\lambda}_+}\right).$$

The power spectrum of noise can be written as follows:

$$S_N(\omega) = \left[ 1 - \left( \frac{Ax_m^2}{2D} \right) \frac{(\bar{\lambda}_- + \bar{\lambda}_+)^2}{(\bar{\lambda}_- + \bar{\lambda}_+)^2 + \omega_0^2} \right] \cdot \frac{(\bar{\lambda}_- + \bar{\lambda}_+)^2 (1+r^2)x_m^2}{(\bar{\lambda}_- + \bar{\lambda}_+)^2 + \omega_0^2}. \quad (10)$$

The output power spectrum is obtained by Fourier transform of the autocorrelation function, as follows:

$$S(\omega) = \frac{\pi}{2} \bar{x}^2(D) [\delta(\omega - \omega_0) + \delta(\omega + \omega_0)] + S_N(\omega). \quad (11)$$

Then SNR is as follows:

$$\begin{aligned} \text{SNR} &= \frac{\int_{-\infty}^{+\infty} S(\omega) - S_N(\omega) d\omega}{S_N(\omega_0)} \\ &= \frac{(1+r^2)\pi A^2 x_m^2 (\bar{\lambda}_- + \bar{\lambda}_+)}{4D^2 [1 - (1/2)(Ax_m/D)^2 ((\bar{\lambda}_- + \bar{\lambda}_+)^2 / ((\bar{\lambda}_- + \bar{\lambda}_+)^2 + \omega_0^2))]}. \end{aligned} \quad (12)$$

For the system,  $x_m = b$  and obtained from formula (2):

$$\bar{\lambda}_- = \frac{k_3^2}{r^2(k_1 - k_2)k_2} e^{-(k_3/D)}, \quad (13)$$

$$\bar{\lambda}_+ = \frac{k_3^2}{(k_1 - k_2)k_2} e^{-(k_3/D)}.$$

If the higher order term of the denominator in formula (12) is not considered, the SNR can be approximated as follows:

$$\text{SNR} = \frac{\pi k_2 (1+r^2)^2 k_3^2 A^2}{4D^2 (k_1 - k_2) r^2} e^{-(k_3/D)}. \quad (14)$$

When  $r = 1$ , the SNR formula of the symmetrical piecewise linear model shown in [5] can be obtained. According to reference [2], the formula of signal-to-noise ratio of continuous bistable system is as follows:

$$\text{SNR} = \frac{\sqrt{2}u^2 A^2}{4D^2} e^{-(u^2/4D)}, \quad (15)$$

where  $u$  is the system parameter, the barrier height is  $\Delta U = (u^2/4)$ , and the well bottom is in  $x = \pm\sqrt{u}$ ,  $u = 2$ . Select the system parameter  $k_1 = 1$ ,  $k_2 = 1$ ,  $k_3 = 1$ . The height of the potential well is the same as the position of the potential well, and then the asymmetry factor is changed to make a comparative study. When the input signal and noise are identical, as shown in Figure 2, it is the change of SNR with the noise intensity  $D$ . It can be seen from Figure 2 that with the increase of the noise intensity  $D$  of the piecewise linear system and the classical continuous bistable system, the SNR also increases rapidly to the maximum value, but the noise continues to increase and decreases. The asymmetric piecewise linear system also has the same phenomenon, but in the process of noise intensity from small to large, the system is always larger than the other two systems. It can be seen that the higher SNR can be obtained for the asymmetric piecewise linear system with the same system parameters.

### 3. Numerical Simulation of Asymmetric Linear Model

In order to verify the above results, the symmetrical piecewise linear model has been compared with the continuous bistable system in reference [5], and the



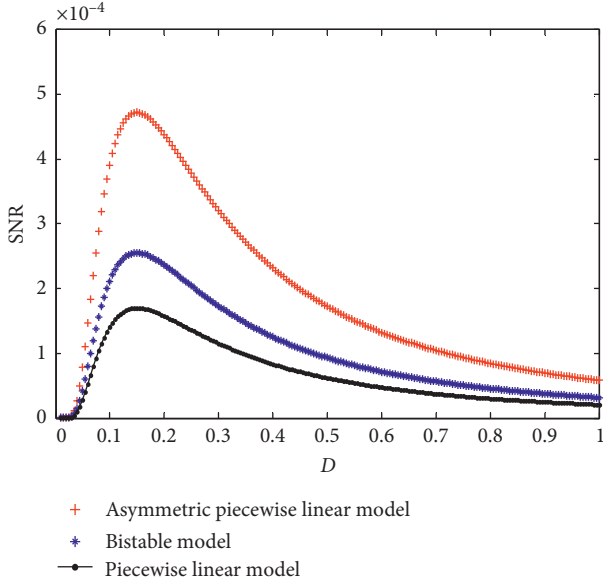


FIGURE 2: Variation of SNR with  $D$  in the three systems.

performance of the piecewise linear model is better. Therefore, only asymmetric and symmetric piecewise linear models are compared. According to formula (1), the system model can be written as follows:

$$\dot{x}(t) = -\frac{dU(x)}{dx} + H(t), \quad (16)$$

where  $H(t) = A \cos(2\pi ft) + \varepsilon(t)$ ,  $\varepsilon(t)$  for noise. Its mean value is zero, and autocorrelation function can be expressed as  $\langle \varepsilon(t) \varepsilon(t + \tau) \rangle = 2D\delta(t - \tau)$ .  $D$  is the noise intensity,  $\tau$  is the delay time,  $f$  is the input signal frequency, and  $A$  is its amplitude. The fourth-order Runge Kutta of (1) is simulated [4], and the resonance behavior of  $r = 1$  and  $r \neq 1$  is compared and explored.

Figure 3 is the input signal frequency  $f = 0.01$ , sampling frequency, and amplitude noise intensity  $fs = 5, A = 0.1, D = 4$ , system parameter  $k_1 = 2, k_2 = 1, k_3 = 0.25$ . Figures 3(b) and 3(d) show the amplitude frequency characteristics of the input signal and the output signal, respectively. It can be seen that the spectrum energy concentrates on the low-frequency component. While Figure 4 is under the condition that other conditions remain unchanged, making the asymmetric factor  $r = 1.5$ , it can be seen from Figure 4(a) that the reduction effect is better than Figure 3(c), and from the frequency domain, it can be seen that the energy concentrated on the input signal frequency is higher.

Figure 5 is a comparison of the average signal-to-noise ratio gain [10] of the asymmetric piecewise linear model and the symmetric piecewise linear model under the same other conditions as Figure 4. It can be seen from the figure that the average signal-to-noise ratio gain of the asymmetric model is always higher than that of the symmetric model with the increase of noise intensity, which is consistent with the conclusion of the formula derivation. Figure 5 shows that the performance of the asymmetric system is not superior to that of the symmetric model because the selection of asymmetric

factors is not optimal. Figure 6 is a three-dimensional graph of noise intensity  $D$ , asymmetric factor  $r$  an average signal-to-noise ratio gain MSNRI under the same conditions as Figure 5. It can be seen that  $r = 2$  is not the optimal case. As you can see, the performance improvement is not very high. This is because the asymmetric coefficient is not very suitable. It is because of this that the adaptive algorithm mentioned below is needed to get the appropriate system parameters.

#### 4. Adaptive Genetic Algorithm

The model proposed in this paper has four parameters, namely  $k_1, k_2, k_3, r$  and the dimension of the genetic algorithm is set as four dimensions [11–13]. The key of the algorithm is to transform the solution into the chromosome needed in a genetic algorithm. There are a variety of “chromosome” transformation methods, which can be divided into real number coding (parameter optimization problem) and integer coding (shortest path problem) according to requirements. The genetic algorithm has a wide range of applications. Usually, some changes will be made according to actual needs. In this study, the range of parameter optimization is determined by theoretical analysis, that is, the range of initial gene in genetic algorithm. By replacing the signal-to-noise ratio with fitness function, the purpose of improvement is achieved. The genetic algorithm can be well applied to the stochastic resonance system. The specific process is shown in Figure 7. Specific steps of the proposed algorithm:

The fitness function of the paper is the output signal-to-noise ratio. Signal-to-noise ratio (SNR) is a common measure in signal processing research. The higher the SNR, the better the fitness. Compared with correlation and average signal-to-noise ratio, the genetic algorithm has lower time complexity and little difference in effect.

Initialize population size and iteration: Random selection of individuals to build an initial population.

The fitness function constructed in step (1) is used to calculate the fitness of all individuals. And keep the best individuals to the next generation.

The individual genes in the population were crossed, and when the crossing process met the variation conditions, the cross was executed (5). The crossover operator of the following formula is used:

$$\begin{aligned} X'_1 &= \lambda_1 X_1 + (1 - \lambda_2) X_2, \\ X'_2 &= \lambda_1 X_2 + (1 - \lambda_2) X_1. \end{aligned} \quad (17)$$

Variation, that is to say, the offspring produced genes that the parents did not have. The construction of the mutation operator is as follows:

$$X' = X + \Delta. \quad (18)$$

Generate offspring and replace any random individual in the offspring with the optimal solution individual in step (3). According to whether the termination condition reaches the maximum number of iterations, the algorithm’s branch flow is determined.

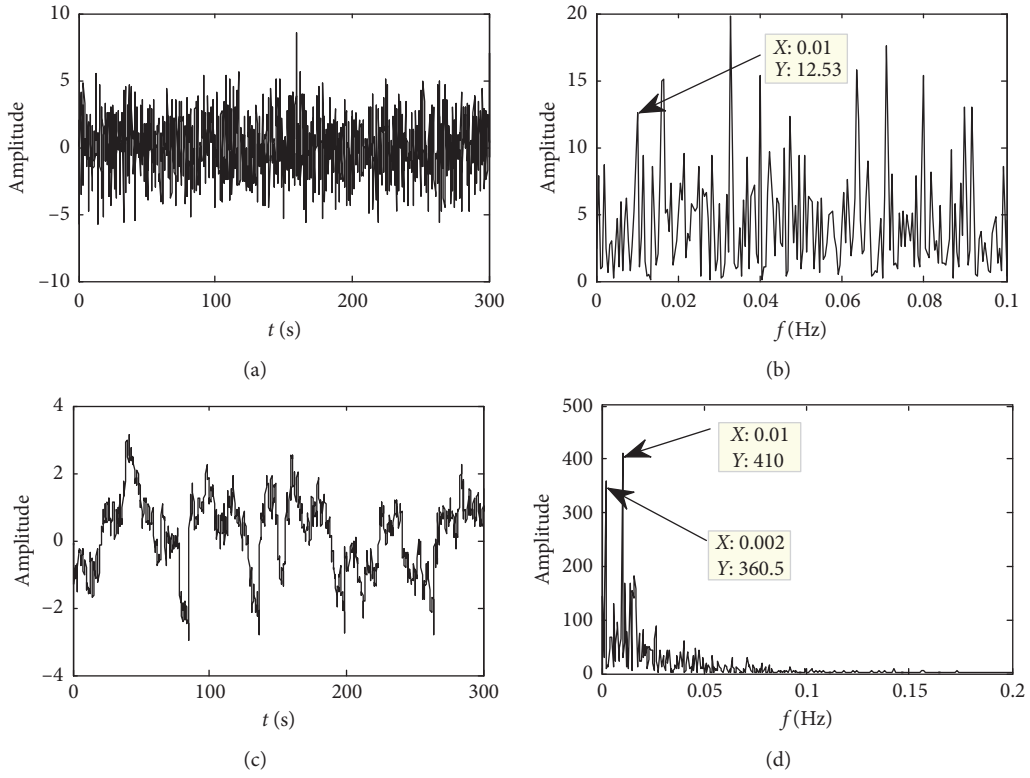


FIGURE 3: Time-domain and frequency-domain graphics of input and output signals.

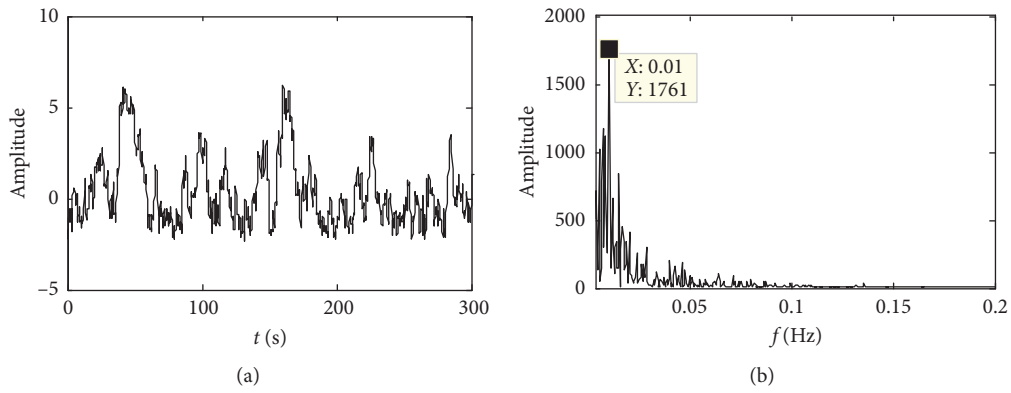


FIGURE 4: Time-domain and frequency-domain graphics of input and output signals.

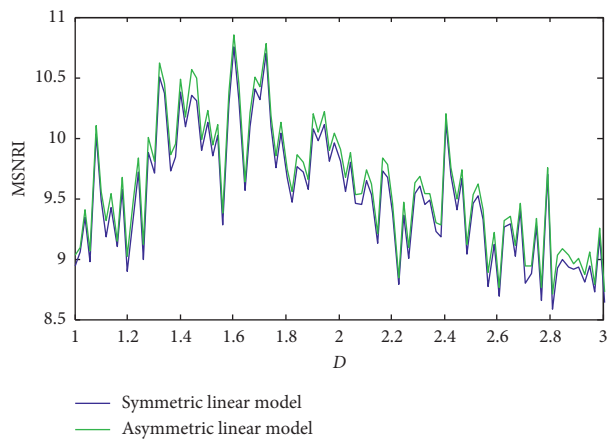


FIGURE 5: Average SNR gain.

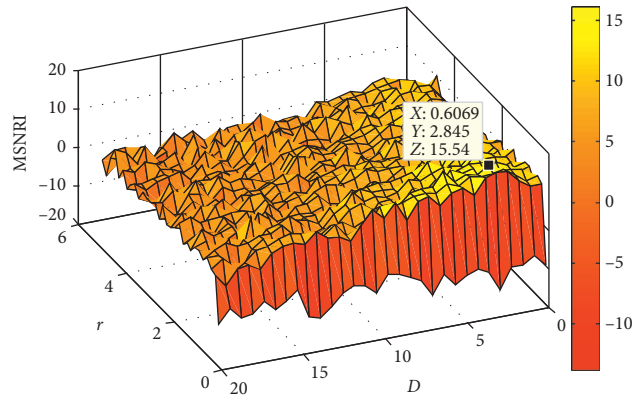


FIGURE 6: Three-dimensional graph of MSNRI varying with  $D$  and  $R$ .

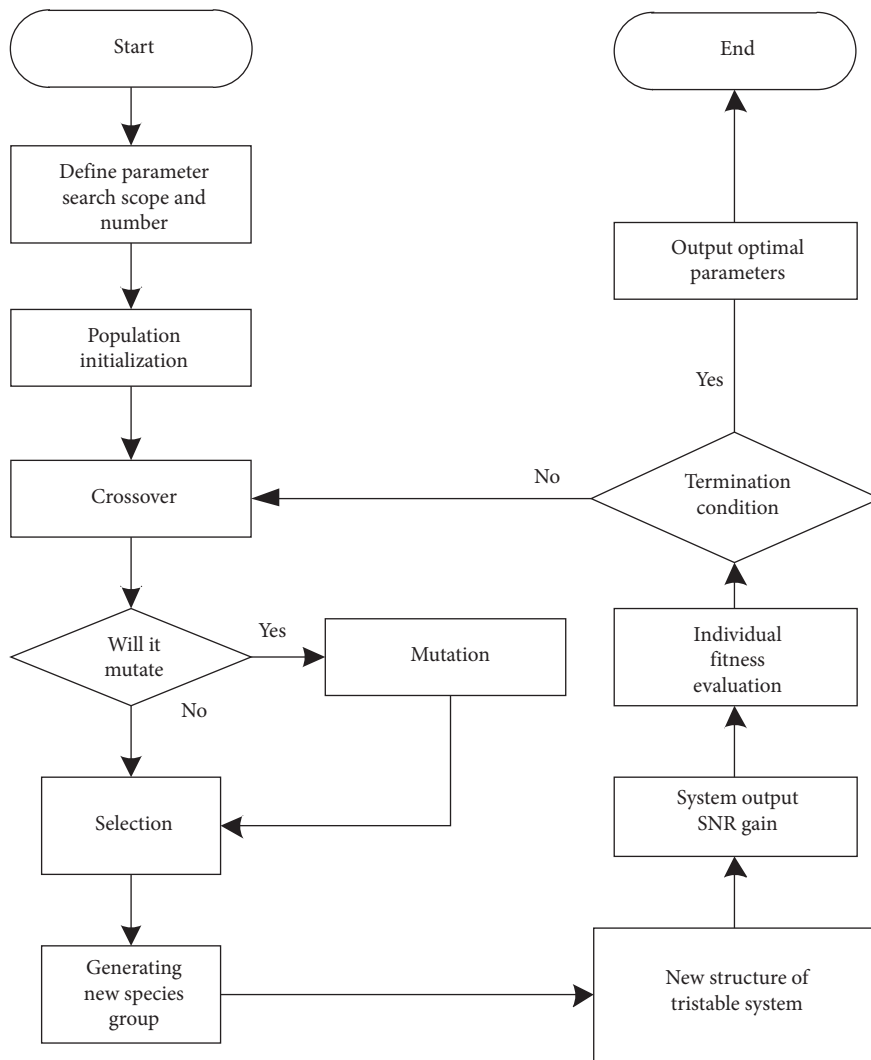


FIGURE 7: Flow chart of genetic algorithm.

Adding optimization parameters to the genetic algorithm will not exponentially increase the time complexity like an ordinary iterative algorithm. The time complexity of the iterative

algorithm is  $o(N^n)$ ,  $N$  is the number of iterations, and  $n$  is the number of parameters. The time complexity of GA is  $o(N * n^n)$ .  $N$  is the genetic algebra,  $n$  is the number of parameters.

TABLE 1: Main parameters of bearing.

| Inner diameter (cm) | Outer ring diameter (cm) | Ball straight (cm) | Thickness (cm) | Stanza (cm) | Number of ball bearings (one) |
|---------------------|--------------------------|--------------------|----------------|-------------|-------------------------------|
| 2.5001              | 5.1999                   | 0.7940             | 1.5001         | 3.904       | 9                             |

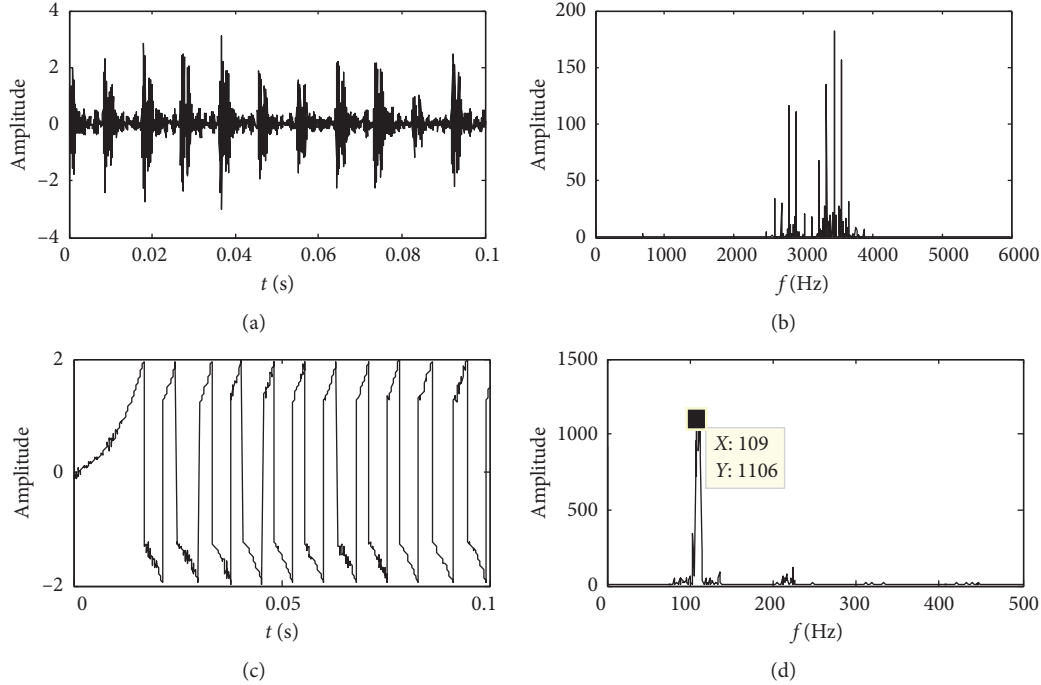


FIGURE 8: Time and frequency-domain diagrams of input and output of outer circle fault detection system ( $a = 0.422000206845763$ ,  $b = 0.01$ ,  $c = 2.645191139445148$ ,  $r = 1$ ).

## 5. Bearing Fault Detection

### 5.1. Fault Characteristics Extraction of the Bearing Type 6205-2RS JEM SKF

**5.1.1. Characteristic Frequency.** The bearing model is 6205-2RS JEM SKF deep groove ball bearing. The main parameters are shown in Table 1. Because the condition of a small parameter is not satisfied, the resonance phenomenon is produced by using the method of second sampling.

Sampling frequency  $f_s = 12000$  Hz, number of sampling points  $N = 10000$ . Here, 5 Hz is selected as the second sampling frequency. The calculation formula of the characteristic frequency is as follows:

$$\begin{aligned} f_{\text{BPFI}} &= \frac{n_r f_r}{2} \left( 1 + \frac{D_1}{D_2} \cos \alpha \right), \\ f_{\text{BPFO}} &= \frac{n_r f_r}{2} \left( 1 - \frac{D_1}{D_2} \cos \alpha \right), \end{aligned} \quad (19)$$

$n_r$  represents the number of rolling elements,  $D_1$  represents the diameter,  $D_2$  represents the bearing, the rotation frequency is  $f_r$ , the contact angle is  $\alpha$ ,  $f_{\text{BPFI}}$  and  $f_{\text{BPFO}}$  are the characteristic frequencies of the inner and outer rings of the bearing, respectively. Substituting the data in the table into equation (19), it is known that the characteristic frequency of

the outer ring fault is 107.28 Hz, and that of the inner ring fault is 162.11 Hz.

**5.1.2. Outer Ring Fault Detection.** Figure 8 shows the input signal of bearing fault to be detected and the output signal of the symmetrical piecewise linear system. It can be seen that the fault signal is completely covered by other high-frequency noises. After passing the SR system, the fault signal is detected. The parameters of this check are all found through the adaptive parameter optimization of the genetic algorithm. The population number of genetic algorithms is 200, the genetic algebra is 200, the crossover probability is set to 0.4, and the mutation probability is set to 0.2. (In the actual natural environment, the crossover probability is much greater than the mutation probability, but in order to jump out of the local optimum, the mutation probability is set to be larger here.)

Figure 9 is an output signal through an asymmetrical piecewise linear system. Compared with Figure 8, it can be seen that the accuracy of amplitude value and inspection at fault frequency is higher than that of the symmetrical segmented system. The time-domain waveforms of the stochastic resonance system are compared. It can be seen that the burr of time-domain waveform of an asymmetric system is obviously less than that of a symmetric system. It is

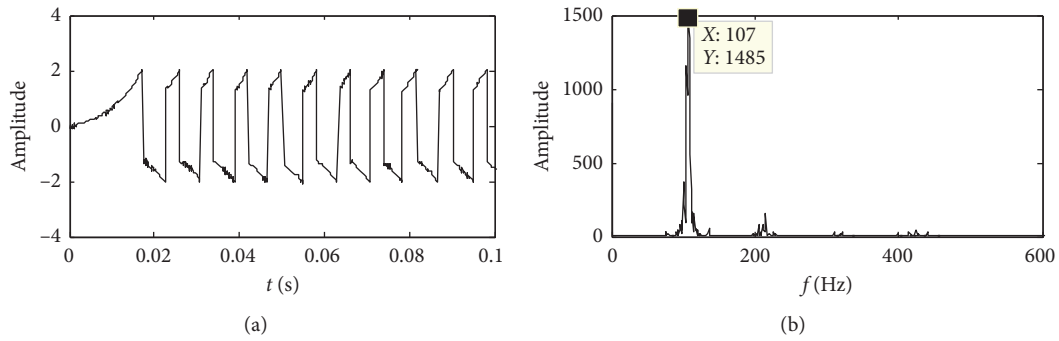


FIGURE 9: Time and frequency-domain diagrams of the output of outer circle fault detection system ( $a = 1.00619166102941, b = 0.11, r = 1.6, c = 0.284256838355532$ ).

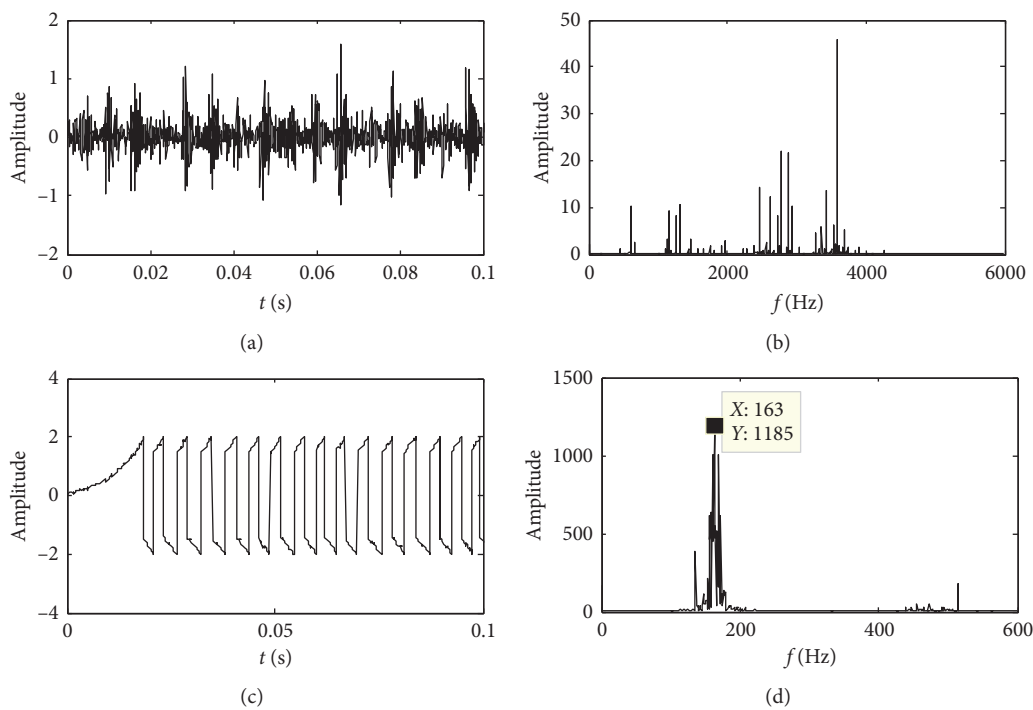


FIGURE 10: Time-domain and frequency-domain diagrams of the input and output of inner-loop fault detection system ( $a = 1.263827550905825, b = 1.226069, r = 1, c = 0.253128458753468$ ).

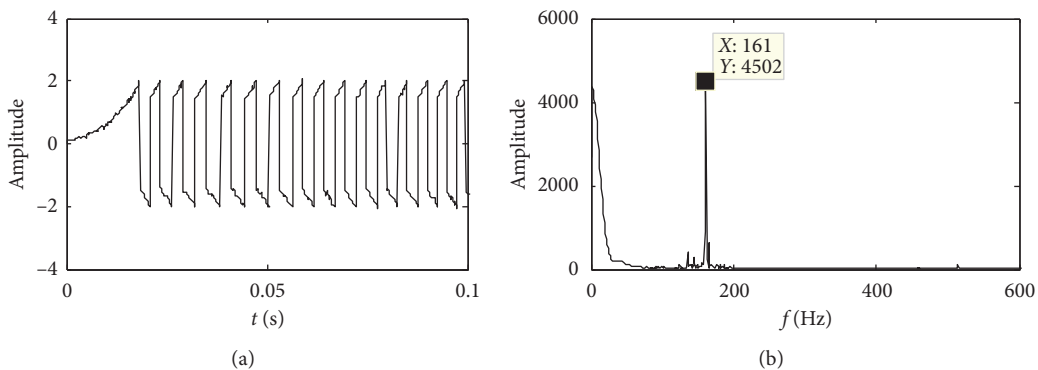


FIGURE 11: Time and frequency-domain diagrams of the output of outer circle fault detection system ( $a = 0.938548936195951, b = 0.11, c = 2.11048612862940, r = 1.6$ ).

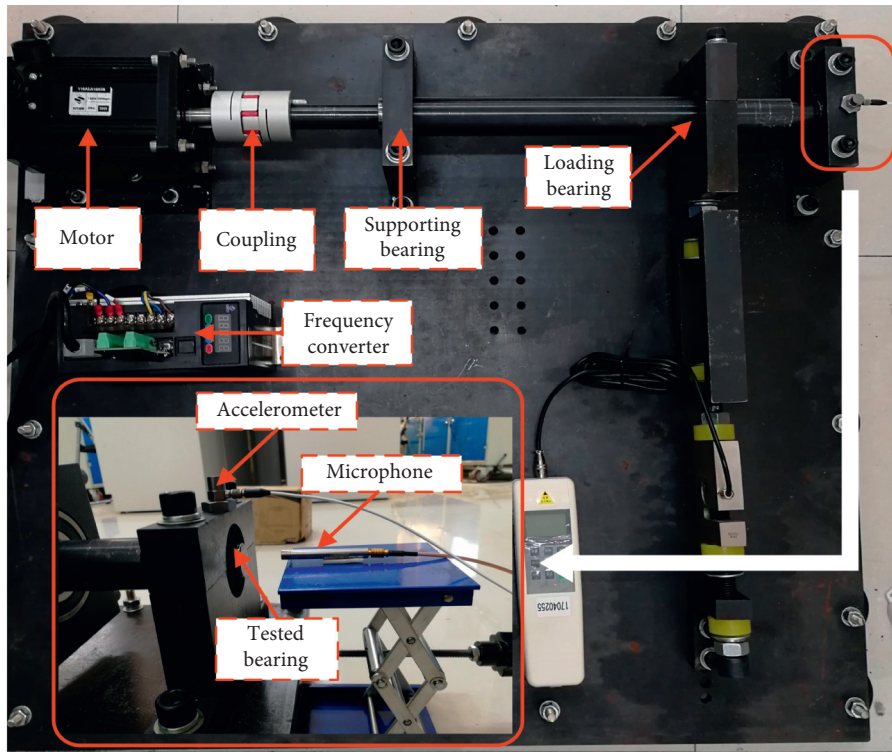


FIGURE 12: The entity diagram of the ID-25/30 test bench.

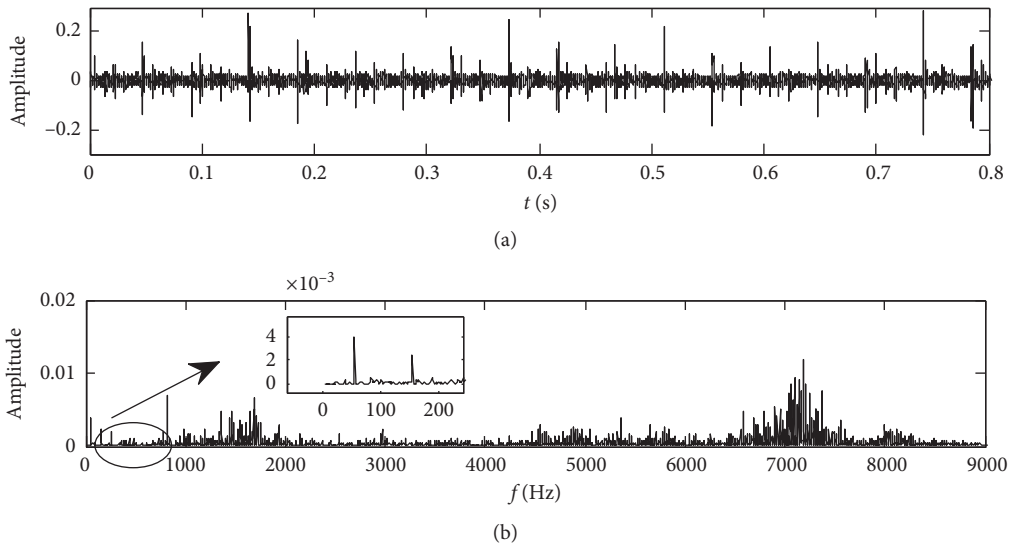


FIGURE 13: Signal to be detected.

consistent with higher amplitude at fault frequency of frequency-domain waveform.

**5.1.3. Inner Ring Fault Detection.** In order to prove that there is still a lot of room to improve the performance of the system, the number of population and the number of iterations are doubled when a genetic algorithm is used for parameter optimization. Figure 10 shows the signal to be

detected for the inner ring fault and the output signal after the piecewise linear system model. It can be found that the fault signal has been detected, but after a longer time of parameter optimization, the system performance has not been significantly improved.

Figure 11 shows the output waveform of the signal to be detected through the asymmetric piecewise linear model. It can be found that the performance is greatly improved after the time of parameter optimization is increased. It is the

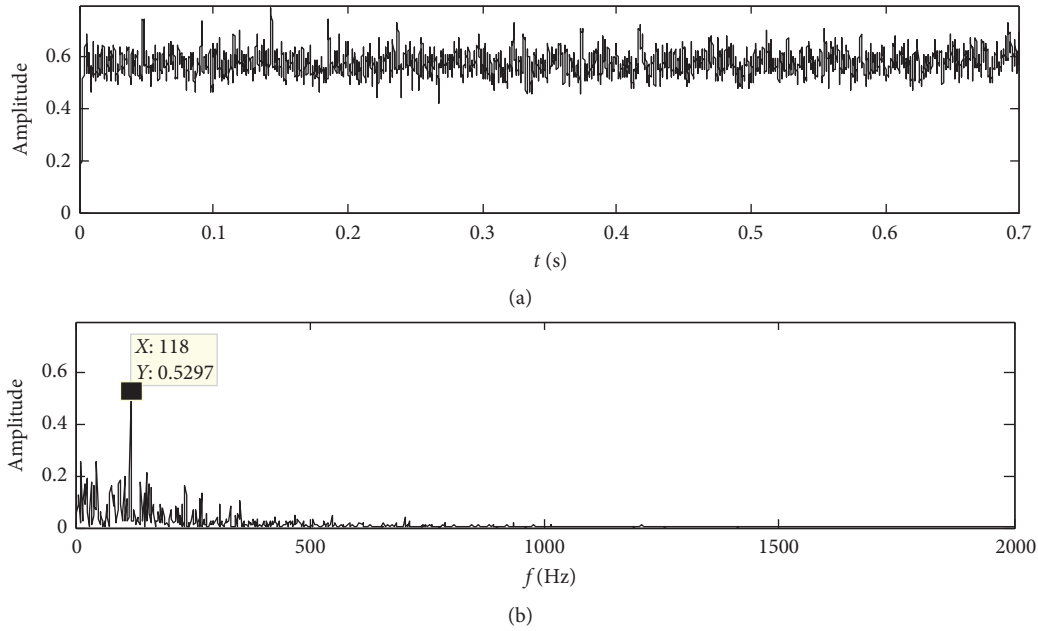


FIGURE 14: Output signal of the piecewise linear system.

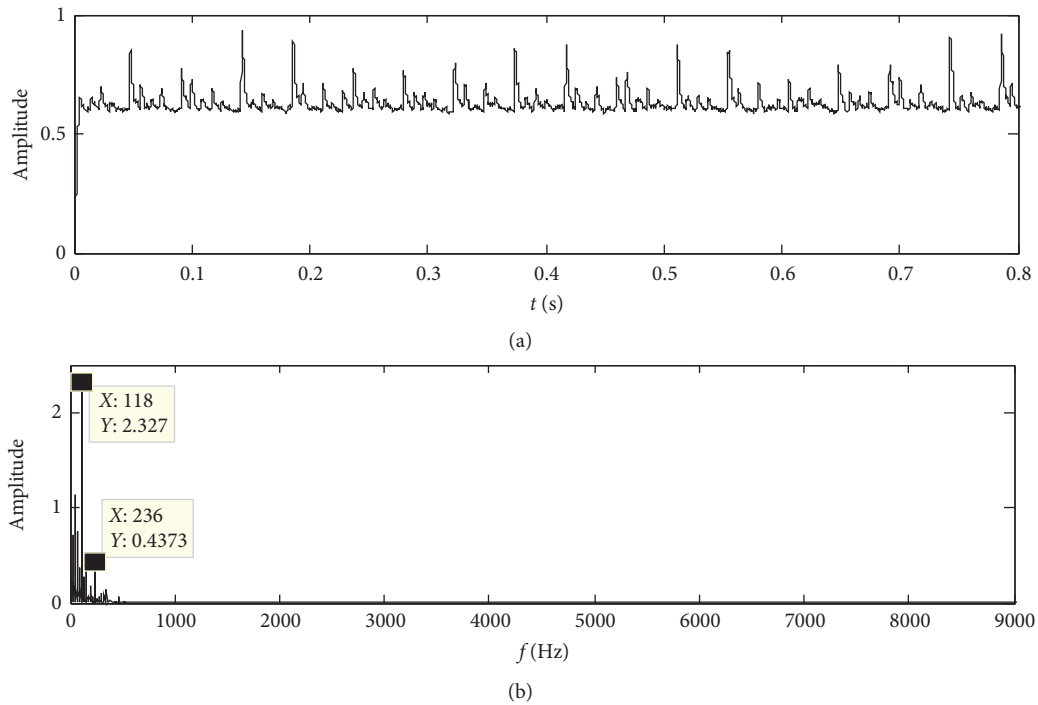


FIGURE 15: Output signal of the asymmetrical piecewise linear system.

same as the previous conclusion and proves the superiority of the system.

**5.2. Fault Diagnosis of ID-25/30 Bearing Health Test Bench.** In order to further verify the practicability of the system, it is applied to another group of experimental platform for verification. The physical figure of the ID-25/30 bearing

health test bench is shown in Figure 12. The calculation method of the characteristic frequency signal of bearing fault is the same as the previous one. The frequency of the inner ring and outer ring can be obtained from the formula. The inner raceway frequency of the bearing can be calculated theoretically to be  $f_{\text{Inner}} = 117.14$  Hz.

Figure 13 shows the fault signal to be tested collected by the platform. It can be seen that the fault model is completely

submerged by noise and cannot be detected. Figure 14 is the output graph of the piecewise linear system. It can be seen that the fault signal is detected. This part of the parameters is also optimized by genetic algorithm, which proves the practical value of the system again. Figure 15 shows the output of the asymmetrical piecewise linear system. Figure 15 has a higher amplitude and the frequency doubling has been checked, indicating that the detection effect performance is very good.

## 6. Conclusion

An unsymmetrical piecewise linear system has good efficiency. This paper discusses the performance of the system from three aspects: formula derivation, numerical simulation, and engineering application. The results show that it has a good performance both in theory and in practice, which is consistent with the formula derivation results. Of course, the performance of a system is closely related to its parameters, and it is difficult to find out the coordination between parameters. Therefore, a genetic algorithm is proposed to find better parameters.

The asymmetry mentioned in this paper is the asymmetry of the width of the potential well and the steady state is assumed to be two steady states, which is not explored in the asymmetry of the depth of the potential well and the multisteady state, so we will continue to study in the following work.

## Data Availability

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

## Conflicts of Interest

The authors declare that there are no conflicts of interest.

## References

- [1] S. L. Lu, Q. B. He, and J. Wang, "A review of stochastic resonance in rotating machine fault detection," *Mechanical Systems and Signal Processing*, vol. 116, 2019.
- [2] S. H. B. Jiao, R. Yang, Q. Zhang et al., "Asymmetric bistable stochastic resonance driven by  $\alpha$ -stabilized noise," *Acta Physica Sinica*, vol. 64, no. 2, pp. 49–57, 2015.
- [3] Y. G. Leng, "Parametric-tuned stochastic resonance mechanism based on Kramers escape rate," *Journal of Physics*, vol. 58, no. 8, pp. 5196–5200, 2009.
- [4] L. L. Tang and G. B. Chen, "Stochastic resonance method driven by well-width asymmetry for fault diagnosis of bearings," *Science and Technology and Engineering*, vol. 18, no. 14, pp. 197–202, 2018.
- [5] L. Z. Wang, W. L. Zhao, and X. Chen, "Theoretical analysis and experimental study of piecewise linear model based on stochastic resonance principle," *Journal of Physics*, vol. 61, no. 16, pp. 50–56, 2012.
- [6] G. Zhang, L. Zhou, and T. Q. Zhang, "Application of improved potential function stochastic resonance in bearing fault detection," *Journal of Electronic Measurement and Instruments*, vol. 32, no. 12, pp. 134–141, 2018.
- [7] L. F. He, X. C. Zhou, G. Zhang, and T. Q. Zhang, "Stochastic resonance analysis of new potential function under levy noise and bearing fault detection," *Vibration and Impact*, vol. 38, no. 12, pp. 53–62, 2019.
- [8] G. Zhang, Y. Shong, and T. Q. Zhang, "Stochastic resonance analysis of exponential monostatic systems driven by levy noise," *Journal of Electronics and Information Science*, vol. 39, no. 4, pp. 893–900, 2017.
- [9] C. Zhang and Y. Y. He, "Bearing fault diagnosis method based on adaptive stochastic resonance and VMD decomposition of genetic algorithm," *Mechanical Transmission*, vol. 42, no. 4, pp. 156–163, 2018.
- [10] H. E. Sun and Z. H. W. Wang, "Weak signal detection of cascade piecewise linear stochastic resonance," *China Mechanical Engineering*, vol. 25, no. 24, pp. 3343–3347, 2014.
- [11] G. Hu, *Stochastic Forces and Nonlinear Systems*, Shanghai Science and Technology Education Press, Shanghai, China, 1994.
- [12] J. G. Feng and I. Sun Shi, "Fourth-order Runge-Kutta method and its application," *Mathematics Learning and Research*, vol. 33, no. 17, pp. 3–5, 2017.
- [13] Z. Zhang, Y. L. Yang, and T. Q. Zhang, "The characteristic analysis of stochastic resonance and bearing fault diagnosis based on NWSG model driven by trichotomous noise," *Chinese Journal of Physics*, vol. 60, pp. 107–121, 2019.



## Research Article

# Decision Tree Algorithm-Based Model and Computer Simulation for Evaluating the Effectiveness of Physical Education in Universities

Zhifei Zhang,<sup>1</sup> Zijian Zhao ,<sup>1</sup> and Doo-Seoung Yeom<sup>2</sup>

<sup>1</sup>School of Physical Education, Zhengzhou University, (School Headquarters), Zhengzhou, Henan 450001, China

<sup>2</sup>Department of Physical Education, College of Arts & Physical Education, Gangneung-Wonju National University, Gangneung-si, Gangwon-do 25457, Republic of Korea

Correspondence should be addressed to Zijian Zhao; [zjzhao@zzu.edu.cn](mailto:zjzhao@zzu.edu.cn)

Received 26 August 2020; Revised 6 October 2020; Accepted 10 October 2020; Published 29 October 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Zhifei Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, the forest algorithm and the decision tree algorithm are mainly used to analyze students' physical education information, course exam results, and student learning data and relevant feature attributes from the online teaching platform. We aim to generate decision trees using the decision tree algorithm for the purpose of generating classification rules, based on which we can find factors that are important to students' physical education performance and form data basis for improving teaching quality to help teaching management and teachers improve teaching methods and adjust teaching strategies. We specifically achieved this objective by constructing a model for assessing the effectiveness of student teaching, the steps of which include data collection and preparation, data preprocessing (data cleaning, conversion, integration), model construction (algorithm training), and algorithm optimization, as well as realizing the simulation results of the model. At the same time, the importance of the relevant attributes of the model is analyzed, and some measures are proposed to improve the universities: the standard of physical education teaching and the corresponding strategies for improving teaching methods. The mainstream development environment is chosen to ensure the complete operation of the project system that integrates learning, operation, and evaluation. The sports virtual simulation experimental teaching system realized in this paper has good functionality, stability, and application benefits in operation and use.

## 1. Introduction

The report "Quality Physical Education: A Guide to Action for Policymakers," published by UNESCO, states that physical education is the only way to combine the development of physical motor skills with the learning and communication of values and is the ideal path for students to acquire core competencies [1]. The study of physical education in universities is the final stage of being able to receive formal physical education and is an important way to refine and improve their knowledge of physical education, and the development of physical education in universities is a prescribed requirement imposed by school education, as well as meeting the needs of the development of sports [2]. The results of the survey on the physical health of

adolescents found that the lung capacity of university students from recent years is a declining trend. Students also have poor vision, neurological failure, and cardiovascular disease, and some of the problems have seriously affected physical and mental health. Compared with university students in developed countries of the world, the fitness and health of university students need to be improved [3]. Existing studies have concentrated on the development of the core competencies of physical education and are directed at transforming and teaching the physical education curriculum, and most of them focus on the theoretical aspects of the basic education field. These studies neglect the factors that influence the development of the person himself as well as other aspects [4]. Therefore, this study is not limited to the disciplinary boundaries but is based on the development of

core human sports literacy as the starting point, in addition to the study of the strategies for the development of core human sports literacy in higher education [5]. Based on the development of higher education, on top of the rich experience in implementing quality education and under the accumulation of practical teaching, the college students' sports core literacy has been promoted [6]. Many factors are influencing the development of college students' physical education core quality. First of all, physical education, as one of the important influencing factors, is the main way for college students to participate in physical exercise, which is closely related to the professional quality of physical education teachers as well as the school's implementation policy, management system, sports environment, and organization of extracurricular sports activities [7]. Therefore, to be able to better cultivate students' sports core qualities, the whole school should be placed in the cultivation system, seek the cultivation implementation way, and optimize the cultivation strategy.

Pai et al. based their PageRank model on the performance statistics and ranking of student-athletes in basketball, hockey, and other sports to assess the individual and team performance of students in a more comprehensive way and to promote their participation in learning [8]. The quantitative performance assessment was used to define the students' game performance, and the qualitative assessment of the students was implemented with a personalized interview form to facilitate the participants' perception and experience of the game performance [9]. At the level of the structural design of the physical education curriculum, Martin et al. in their paper analyzed the new concept of physical education curriculum in comparison with traditional teaching [10]. On this basis, they discuss the importance of classroom assessment as it evolves. The research points out that the New Physical Education (NPE) curriculum perspective, based on national development and youth health issues, has significant implications for traditional curriculum design, implementation, and assessment [11]. Furthermore, research on assessment for physical education and learning has focused on communication and linkages between schools, communities, and families; for example, Kim et al. studied parents' perceptions of school assessment and the need to improve the school's curriculum based on parents' views [12]. Dieter et al. have further analyzed the physical education model (SEM) to assess students' motivation from the perspective of self-determination theory, stating that students need a certain time to determine their behavior in an educational environment. One of the most effective means of dealing with classroom assessment is the design and use of rubrics [13]. Mao et al. state that the use of rubrics to determine students' overall learning performance can greatly enhance students' motivation [14].

As can be seen, there are quite several articles and writings on assessment for physical education learning. Whether it is about the relationship between learning assessment and curriculum development or the specific implementation and application of learning assessment, it is evident that physical education learning assessment has become an important hotspot in today's education sector. It

is necessary to establish the educational concept of "health first" and to provide adequate physical education classes, to help students enjoy fun, strengthen their physiques, enhance their personalities, and temper their wills in physical exercise. These important discourses have greatly elevated the status of college sports and are important guidelines for college personnel training. Apply the software theory to the evaluation system functional requirement analysis, design the main functional modules, realize the decision support system based on data mining and simulation, improve the work efficiency of the evaluation department, and provide decision guidance through certain mining analysis. The research proposes a simple and effective data warehouse design and implementation method and uses online analysis and processing technology to analyze the data in the data warehouse and provide basic data for subsequent data mining. After studying and researching the commonly used data mining algorithms and understanding the characteristics of different algorithms, a decision tree algorithm suitable for this data mining is selected based on the data studied in this paper, an improved algorithm is used to establish the mining model, and the mining results are analyzed to extract valuable decision information. According to the problems found by applying the decision tree algorithm, the algorithm is studied in depth, analyzed, and compared, combined with the characteristics of the data of this project, and the analysis of data mining and data warehouse and online analysis and processing techniques are used to obtain useful information to provide a meaningful reference for decision making in the evaluation of physical education.

## 2. Design of a Decision Tree-Based Model for Assessing Teaching Effectiveness in Universities

*2.1. Improved Decision Tree Algorithm and Computer Simulation Design.* As one of the most commonly used methods for data mining, the decision tree algorithm has been widely used in different fields since its introduction. It has undergone a long process of going from shallow to deep and from simple to complex. The decision tree is a top-down, recursive division that uses a top-down, divide-and-conquer approach, and its basic algorithm is essentially greedy [15]. Starting from the root node, each nonleaf node is found to find an attribute in its corresponding sample set to test the sample set, and the training sample set is divided into several subsamples according to the different results of the test. Each subsample set constitutes a new leaf node, and the above process is repeated for the new leaf node so that the loop continues to reach a specific termination condition [16]. The flexible approach to management adopted by teachers and teaching, which is determined by the characteristics of university work, is to go to class when there are classes and then to have free time when there are no classes. While this is conducive to research and learning, it is also likely to keep certain aspects of teaching, such as lesson planning, homework revision, and lab instruction, out of control. The

decision tree adopts a top-down recursive approach to compare and evaluate the attribute values of nodes within the decision tree and determine the branch down from the node based on the different attribute values. One of the biggest advantages of the decision tree-based learning algorithm is that it does not require the user to acquire a lot of basic knowledge during the learning process, as shown in Figure 1.

The process of constructing a decision tree is divided into two steps: tree building and pruning. The first step is the tree building stage, which selects part of the training data and builds a decision tree by the breadth-first recursive algorithm until each leaf node belongs to the same class. The second step is the pruning stage, which uses the remaining data to check the generated decision tree and correct the errors, and it finally prunes the decision tree and adds nodes until a correct decision tree is built. The decision tree building algorithm is a recursive process that ultimately results in a decision tree, and pruning reduces the impact of noisy data on classification accuracy. In general, the greater the information gain, the greater the “purity improvement” obtained by using features to partition the dataset. Therefore, information gain can be used to select attributes for decision tree partitioning, which is to select the attribute with the greatest information gain.

Information gain is the amount of change in the desired information, which mainly reflects the method of sample feature importance, and there is a positive correlation between the two: the importance of the sample feature increases and decreases with the amount of information gain. Information entropy is mainly used to measure the amount of information, so we define the entropy  $w$  of data [17].

$$\begin{aligned} Q_L(w) &= \frac{1}{2} \sum_{i,j} P_{ij} (w^T x_i - w^T x_j)^2 \\ &= \frac{1}{2} \sum_{i,j} P_{ij} (w^T x_i - w^T x_j)(w^T x_i - w^T x_j). \end{aligned} \quad (1)$$

The information gain rate and destination information gain will tend to select the attribute with the larger information gain value. It applies a parameter value to the specification of the information gain, which is calculated as follows [18].

$${}^G D_t^v f(t) = \lim_{h \rightarrow 0} \frac{1}{h^v} \sum_{m=0}^{\lfloor t-a/h \rfloor} (-1)^m \frac{\Gamma(v+1)}{m! \Gamma(v-m+1)} f(t-mh), \quad (2)$$

and based on this, the formula for calculating the gain rate can be obtained as follows:

$$m_i(g) = \frac{\text{fit}_i - \text{worst}(g)}{\text{best}(g) - \text{worst}(g)}. \quad (3)$$

There are many algorithms for generating decision trees in data mining, and several typical decision tree generation algorithms are highlighted below. The ID3 algorithm is the most influential and typical in decision tree mining, which

selects test attributes for each nonleaf node in the tree through a set of rules based on information theory, using entropy as the basis for classification, and finally classifies the data into the form of a decision tree [18]. The basic idea of the ID3 algorithm is to use information entropy as a measure of attribute selection of decision tree nodes; each time it first selects the attribute with the most information, i.e., the attribute that can change the entropy value to the smallest, to construct a decision tree with the fastest decrease in entropy, so that the entropy value at the leaf node is 0 [19]. Currently, the set of instances corresponding to each leaf node belongs to the same class. Since the attribute with the highest gain of information is always selected to divide the rules, the algorithm classifies data quickly, the depth of the tree is average, and the division rules are simple. The ID3 algorithm is important in the process of constructing a decision tree: attribute determination and set partitioning. The process of attribute determination is to select the attribute with the largest value of gain information from the selected data as the root node and construct a branch with the attribute value of this node to divide the selected data into several unrelated subsets; after branching the nonleaf node, it will determine the attribute of the subset value twice and then continue to branch until it branches to the leaf node.

C4.5 is a classification decision tree algorithm in machine learning, which is based on the ID3 algorithm and improves on the advantages of ID3 again. This algorithm uses the concept of information gain or entropy reduction to select the optimal division, to better realize the construction of the decision tree: (1) It selects attribute columns based on the information gain rate, overcoming the deficiency of biased selection of attributes with more values when selecting attributes with information gain. (2) It constructs a decision tree. The tree pruning process is completed in the tree construction process. (3) The data of continuous attributes are processed by applying discretization. (4) The tree pruning construction is also used for incomplete data. The ID3 algorithm can only solve the discrete data attributes, but the C4.5 algorithm can handle the discretized continuous attributes well, with the following procedure.

$$Q(w) = \frac{Q_N(w)}{Q_L(w)} = \frac{\text{tr}(w^T X L_N X^T w)}{\text{tr}(w^T X L X^T w)}. \quad (4)$$

Computer simulation (data visualization), as the name implies, is a simulated representation of data that uses computer graphics and image processing techniques to interpret data through analysis, transformation, and graphic patterns (including animation) in either planar or three-dimensional form, while providing methods, theories, and techniques for interaction [20]. It helps people to view data and the relationships between data more visually. The atlas of computer simulation is composed of a large amount of data. Each data item is the basic unit of the atlas, and the value of each attribute of the data is represented by the multidimensional terms so that the user can observe and analyze the data from different angles to better serve the user and help the user to make decisions. Simulation technology is the data mining results of abstract information, with a

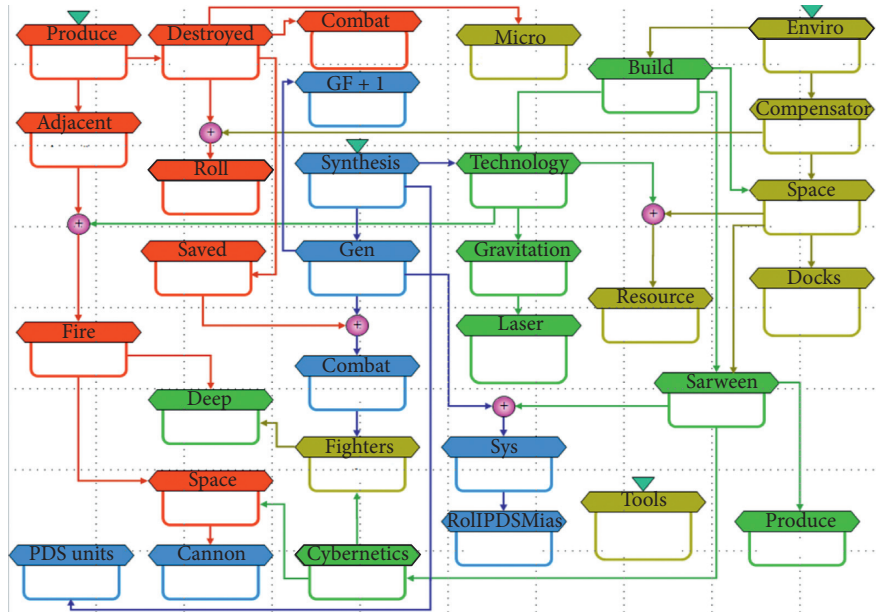


FIGURE 1: Model of improved decision tree algorithm.

simple and intuitive form of expression, to deepen the user's understanding of the meaning of the data, while understanding the interrelationship between the data and development trends. The characteristics of computer simulation technology are as follows. There is strong interaction with the user. The user is no longer simply a receiver of information but can also manage, process, and develop the data. Computer simulation technology can classify, arrange, and present data from multiple perspectives. For example, users can view data time, percentage, rank, and other dimensions. Users can view data and analyze it using simulation charts, histograms, line graphs, puzzles, etc. [21].

Scientific computational simulation refers to the use of computer graphics and image processing techniques to display engineering measurement data, data generated by scientific calculations, and calculation results on the screen and transform them into images. They can also be processed interactively using theoretical methods and techniques. The classification of scientific computing data is relatively extensive and can be divided into structured, unstructured, and mixed data according to different data structures or scalar, vector, and tensor according to data types. Scientific computing simulation technology faces two difficulties in the development process: first, the way to classify the object data by research; second, how to display the simulation object on the screen practically and effectively, making users view it interactively.

*2.2. Evaluation Design of the Effect of Physical Education in Universities.* Model building is at the heart of data mining, which means determining which algorithm to choose to build the decision tree. The implementation of the algorithm requires certain tools for programming. In this paper, MATLAB tools are used to build the CART decision tree. There are three main reasons for this. Firstly, the research

content of data mining is the learning and application of models and algorithms, and MATLAB is particularly suitable for algorithm development. Because MATLAB can directly call a rich library of mathematical functions to quickly implement algorithms according to the flow steps of the algorithm, the programming workload using MATLAB will be greatly reduced when the computational requirements are the same. The syntax of MATLAB is more intuitive than Python or R, and the learning and programming of algorithms are easier. Secondly, MATLAB has efficient and rich scientific computation functions, including calculus, matrix computation, and symbolic computation, and system simulation applications are very extensive. Thirdly, MATLAB itself is a program development tool with a friendly GUI development function [22]. MATLAB was used to learn the principle of algorithms in depth, and then hands-on programming computer programs were deployed to solve many machine learning problems in practical applications. Using MATLAB is more focused, and the toolbox is mostly developed in C. The tools are all interconnected, and it is easier to get started if you have or understand this data mining knowledge and then apply these tools, as shown in Figure 2.

From the above rules, by analyzing the meaning of the rules in detail, some teaching strategies can be adjusted for the problems that arise, such as the need to have relatively effective teaching methods in the online teaching platform, whether to make it mandatory or set open assignments (conscientious students do open assignments, while lazy students do mandatory assignments) and guide students to learn independently. Other aspects, such as out-of-province students with a poor foundation and students with a good foundation who also fail exams, can be predicted pre-intervention can be made; students with good English scores generally have a positive attitude to learning and are encouraged to study professionally. Since the data is not

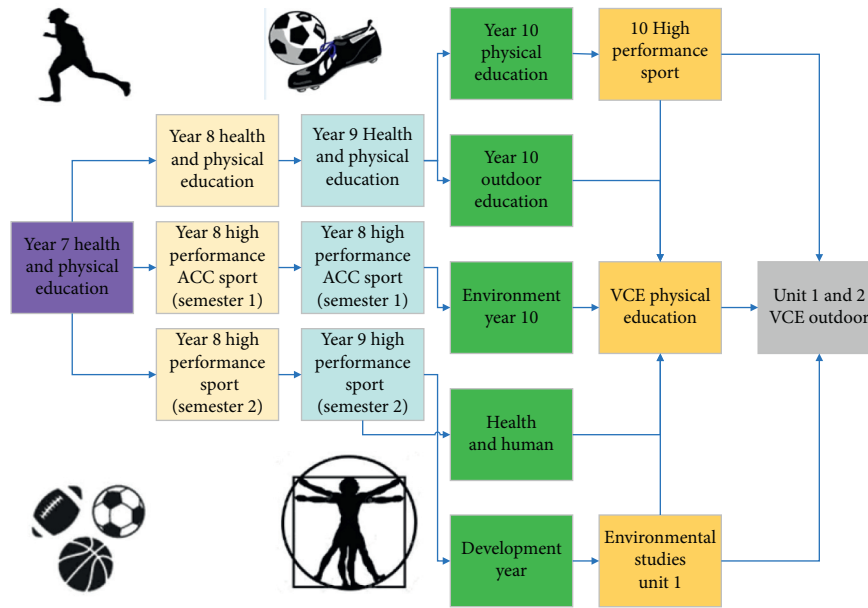


FIGURE 2: The design of physical education effectiveness evaluation.

complete and comprehensive, the results obtained can only serve the major students' local construction advice, but to some extent to make the assessment of student learning effectiveness possible. With the continuous deepening of the new curriculum, curriculum reform is steadily advancing in depth; in the gradual integration of the new curriculum concept and classroom teaching, we have made some achievements and accumulated some experience in the classroom. At the same time, many problems often arise in the classroom, which we did not anticipate in advance, when we must think of a good strategy and find a good way to solve them.

With the development of colleges and universities and the advancement of network technology, the introduction of data mining technology, visualization technology, etc. into the assessment information management system of colleges and universities will provide great convenience and work efficiency. Besides, through data mining technology to analyze student information over the years, you can find some valuable laws and factors that affect the assessment of student reporting, hence more scientific guidance to adjust our plans and decisions. The design of the database system directly affects the quality and operation of the assessment management system. The relative independence, data integrity, and consistency of the database were followed in the design of the database. According to the design features of this system, the traditional database processing mechanism differs greatly from the decision analysis needs and cannot handle online analysis processing, so a data warehouse was established based on the database for the data between different databases. The evaluation management information of this system uses SQL Server 2012 as the software for data storage and management, and this paper only takes the design of the evaluation database as an example, aiming to illustrate the relationship between data tables, databases, and data mining functions and visualization functions.

**2.3. Performance Evaluation Design.** When performing data mining algorithms, to improve the college's reporting success rate, the assessment dataset is used as the object of study; therefore, the decision tree algorithm is used to analyze the key attributes of the assessment data source that may affect the reporting rate and to identify the factors that are most likely to affect the student reporting, so that these potentially relevant influences can be scientifically applied to provide key decisions for future assessment work. The advantages of decision tree algorithms include easy-to-understand analysis, high classification accuracy, and high execution efficiency, making decision tree algorithms well-suited for use in mining large amounts of data. The commonly used decision tree algorithms are ID3 and C4.5, but the two algorithms are very different in the selection of splitting rules. ID3 selects the classification criterion of information gain, which is to select more attribute values as splitting attributes, and it can only mine nonlinear data. C4.5 evaluation classification criterion is the information gain rate, that is, according to information; therefore, C4.5 can effectively circumvent the problems of ID3. Moreover, through the comparative analysis in Section 2.1, the decision tree algorithm of C4.5 is preferred to be used for data mining for better attribute factors, as shown in Figure 3.

The improved algorithm merges branches with higher and lower entropy values, i.e., branches where the division is not important, effectively reducing the fragmentation problem, improving branch efficiency, and limiting the impact of the overfitting problem. In contrast, this paper studies the promotion of assessment in universities, with diverse student sources, but the number of people assessed in each major is relatively small due to local and other reasons, so this requires a more balanced analysis of the diverse assessment channels and student sources, to prevent analysis bias. The improved C4.5 decision tree algorithm makes better use of memory space, balances the selection of

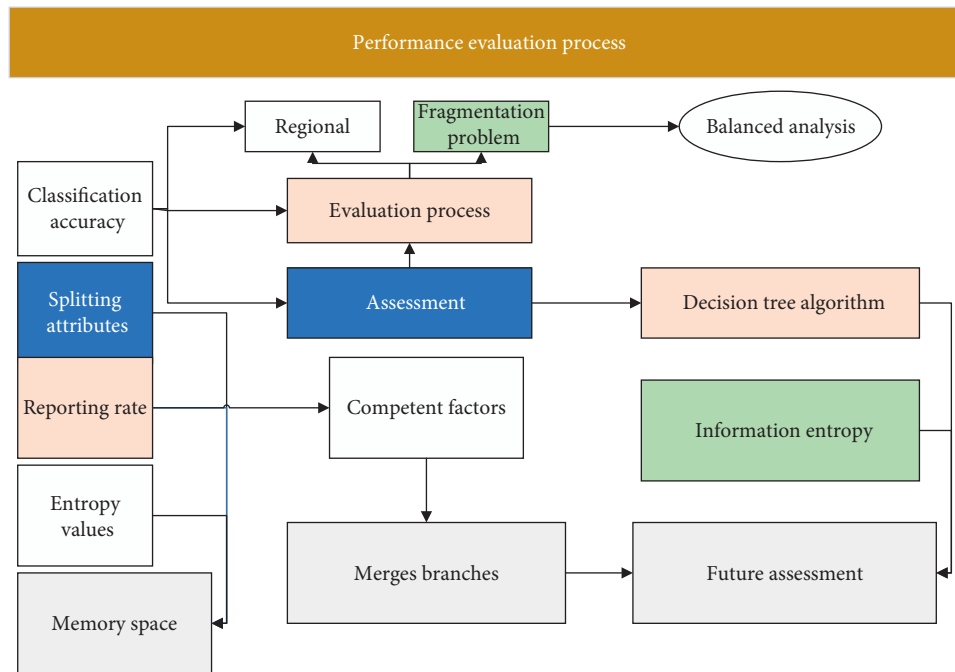


FIGURE 3: Performance evaluation process design.

information entropy, and avoids low or high information entropy caused by the human influence of competent factors, which is more conducive to a more accurate decision tree.

Due to the small data amount selected in the selection of the training set, there may be some deviations in the obtained mining model, and then the training set model is verified through the test set to verify whether the model is accurate, so that the mining model can be further improved and revised, which will be analyzed in the decision tree analysis function module of the system module analysis.

### 3. Results and Analysis

**3.1. Analysis of Assessment Results.** It is found that the more the decision tree base classifiers are, the smaller the generalization error will be. The accuracy of the algorithm is close to 98.25% compared to 8258% for a single base classifier. This is comparable to the accuracy of the CART algorithm. This shows that the decision tree algorithm can predict new data better than the CART algorithm.

In Figure 4, it is found that the prediction error rate of the classifiers is decreasing; i.e., the prediction accuracy rate is increasing. It is more clearly known that the algorithmic accuracy of the decision tree keeps improving with the increase in the number of classifiers. Evaluating the decision tree classification performance cannot be done by simply counting the accuracy of a single breakpoint, but it requires analysis of the classifier's ROC curve, AUC value, and other metrics. The decision tree algorithm is used to validate predictions for 30.25% of the test sample and compare them with the true category. The ROC curve and AUC value are calculated and Figure 5 is obtained. It has an AUC value of

0.7444, although the ROC and AUC metrics are improved by a point over the CART algorithm and the decision tree is more accurate. However, whether the comprehensive performance of this algorithm can still be improved remains to be explored further by the experiment. The ROC (Receiver Operating Characteristic) curve and AUC (Area under the Curve) values are often used to evaluate a binary classifier (the advantages and disadvantages of ROC and AUC (classifier) are discussed here). The ROC and AUC are used in the evaluation of model prediction results when doing medical image computer-aided pulmonary nodule detection in the past. The characteristics of ROC and AUC are briefly introduced here, and how to make ROC plots and calculate AUC values is discussed in more detail.

An analysis of the unsatisfactory experimental results, in terms of the distribution of the dataset, reveals possible causes. One is the problem of data imbalance in the dataset. The target variable classification ratio in the two-category dataset was 586:110 (pass or fail). Afterward, we adjusted the target classification ratio to 407:258 (mean division) for training, made predictions on 30.25% of the test data, and then conducted the ROC curve and AUC value statistics. The expected experimental effect was achieved. The decision tree algorithm can predict the classification more accurately. However, the importance of the attributes of the independent variables cannot be represented as clearly as the decision tree. According to the most important attribute summarized in the decision tree (professional grades), after removing this attribute and retraining the classifier, the false and true rates of the test set can be calculated to obtain the ROC curve Figure 6, which has an AUC value of 0.611 and a decrease of 23.34%, which shows the importance of this attribute. Similarly, the ROC curve and AUC value can be

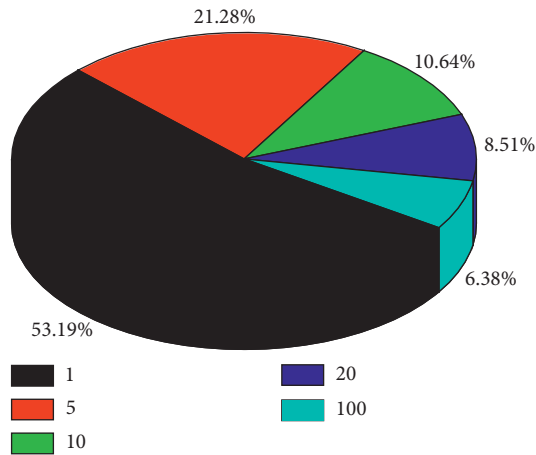


FIGURE 4: Number of base classifiers and generalization error.

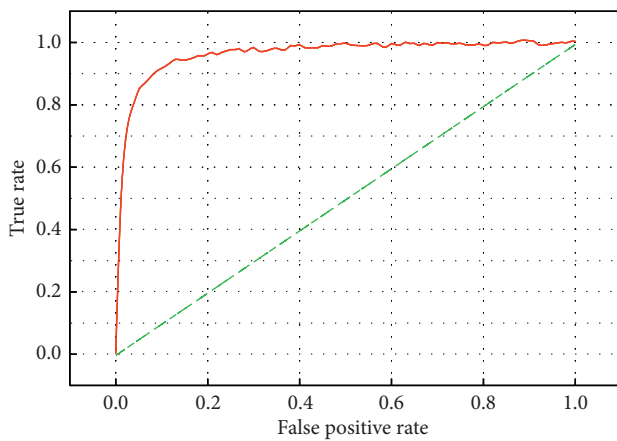


FIGURE 5: ROC of a decision tree on 30% test set.

statistically determined by removing the other attributes. In this way, we can indirectly analyze the characteristics of the most important attributes that affect the target variable.

The AUC of the highest decision tree algorithm is only 0.8441, which is far from excellent performance. The highest decision tree algorithm has an AUC of only 0.8441, which is far from excellent performance. However, the actual classification application in this dataset can predict the classification more accurately and meet the practical application requirements.

The CART and RF algorithms are used to train the data. Firstly, the CART model is used as a single classifier, and the CART classifier is used as a base classifier. The data is trained using the random forest algorithm with different numbers of base classifiers, resulting in four integrated classifiers and a total of five classifiers. The new data table “data” is predicted using each classification model, and the predicted results are compared with the actual classification labels, as shown in Figure 7.

The results show that for both decision tree (CART) and random forest (RF) algorithms, by constructing a random forest algorithm based on CART classifier, the random forest

algorithm integrated with a different number of base classifiers enhances the classification evaluation indexes, in which the true rate, precision-P, recall-R, and F1 value are improved, and the classification performance of the model reaches the expected results. The accuracy of the random forest algorithm is 0.981 for the 5-tree classification model, 0.99 for the 10-tree classification model, and 0.991 for the 20-tree and 100-tree classification models. The classification accuracy has reached the ideal state. As for the comprehensive performance ROC and AUC metrics of the classifier, a comparison has been given in the previous section and will not be repeated here.

In the process of predicting student achievement using classification algorithms, the performance of the final classification model is related to not only the criteria of algorithm goodness and split attribute selection, but also the collection of raw dataset and data preprocessing. The key to classifying and predicting student performance is the need to filter out the main attributes that influence student course performance from the attribute characteristics of the raw data, use this influencing factor as a candidate attribute set, and construct a classification prediction model with student test scores as the target variable. It is found that the attribute set of the raw data may not be very comprehensive and may lack some of the main feature attributes; then, the accuracy of the constructed classification model will not be very high. Therefore, the algorithm is not a panacea; there is no perfect algorithm in the world, and perfect data needs to be collected to implement it. Anomalous data in the dataset can also affect the classification accuracy of the model. It affects a single decision tree model more, while integrated combinatorial models are relatively less easy to affect, and it has a strong generalization ability. However, if multiple trees are trained on the same data, it is also easy to get strongly correlated trees, and then the effect is rather bad, which is all related to the choice of the sample set and attribute selection.

**3.2. Analysis of Computer Simulation Results.** According to the system test case design scheme, the virtual simulation experimental system of “Human Movement Ability Assessment and Fitness Path Design” was tested. 48 test cases were set up and the test report is shown in Table 1.

Among them, 18 user interface test cases, 20 functional test cases, and 10 performance test cases passed the test at a rate of 100%, which shows that the system has passed the application test and can be put into teaching application, as shown in Figure 8.

The interaction of the system is the basic characteristic of virtual simulation experiments. According to the constructivist theory, keeping good interaction in the learning process can effectively stimulate learners’ motivation and enhance the learning effect. Ease of use is an important dimension to evaluate the logic and science of the virtual simulation experimental teaching system in terms of functional design, which can provide important instructions for the design and optimization of the system. Besides, the vivid experimental scenarios and easy-to-use operation

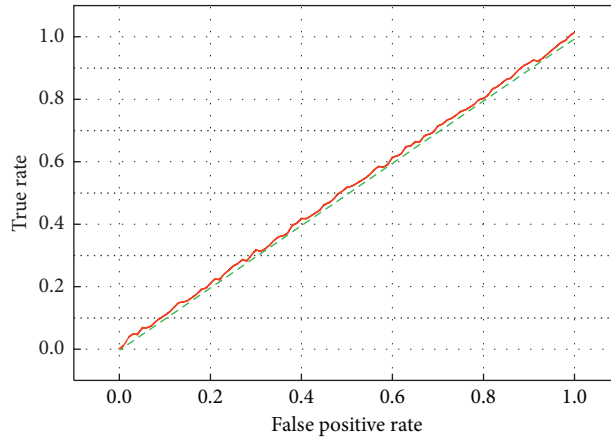


FIGURE 6: ROC curve with the main independent variables removed.

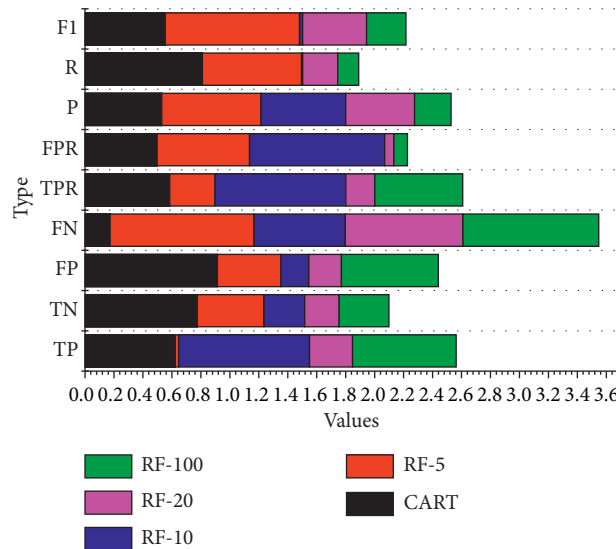


FIGURE 7: Performance comparison of CART and RF classifiers.

TABLE 1: Test cases.

| Test cases | Number of questionnaires issued | Number of questionnaires withdrawn | Number of valid questionnaires | Recovery rate (%) | Efficiency (%) |
|------------|---------------------------------|------------------------------------|--------------------------------|-------------------|----------------|
| Numbers    | 12                              | 16                                 | 20                             | 93.65             | 94.86          |
| Parents    | 14                              | 14                                 | 20                             | 87.47             | 86.74          |
| Total      | 14                              | 15                                 | 19                             | 88.47             | 85.78          |

procedures are also beneficial to the efficient operation of the knowledge learning process, which is an effective extension of the traditional experimental sports education.

The usefulness of the system is one of the indicators to evaluate the teaching effect of sports virtual simulation experiments. The goal of sports virtual simulation experiments is to meet the actual needs of sports experimental teaching and improve the teaching effect of the experiments. By evaluating the usefulness of the system, we can clearly understand the learners' knowledge of virtual

simulation experiments, to provide a basis for the assessment of the practical effect and teaching significance of sports virtual simulation experimental teaching system. The motivation of the system is the psychological feedback of the learners after using the virtual simulation experimental system. The high level of motivation indicates that the learners recognize the value of the virtual simulation experiments in terms of learning effect; on the other hand, it also indicates that the system can dig deep into the learners' learning needs to strengthen their desire for



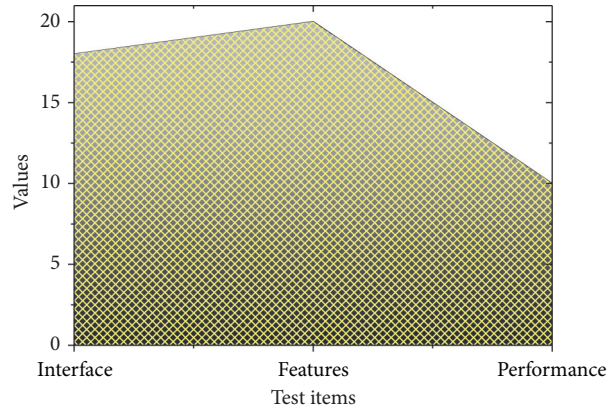


FIGURE 8: Test report of virtual simulation experimental system.

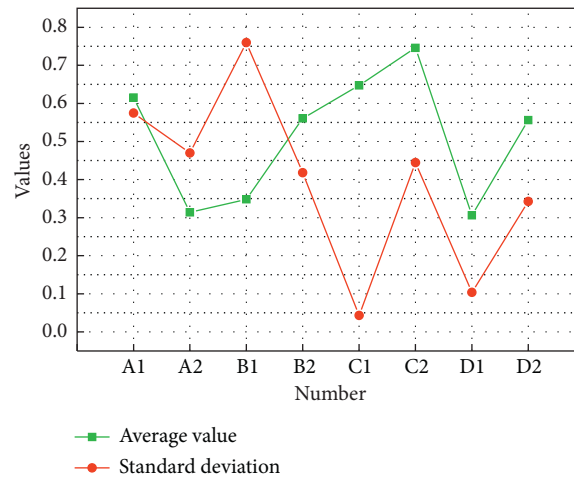


FIGURE 9: Descriptive statistics of sports virtual simulation experiment.

learning sports knowledge, which reflects the guiding role of the virtual simulation experiments in sports to the learners. The development of virtual simulation experimental teaching in higher education is an important content of higher education informatization. This paper points out the necessity of virtual simulation experimental teaching according to the experimental practice of higher education in China, analyzes the problems of the current virtual simulation experimental teaching project, and proposes the sustainable development mechanism of virtual simulation experimental teaching project.

The application function and teaching effect of sports virtual simulation experimental teaching are evaluated by a Likert five-point scale and the specific results of statistical analysis are shown in Figure 9. The above results show that the test subjects are highly satisfied with the virtual experiment of “Human Movement Ability Assessment and Fitness Path Design,” with an average score of more than 4, which indicates that the four dimensions of interactivity, usability, usability,

and motivation of the system are at a relatively good level, and the system is recognized and praised by the test subjects.

From the evaluation feedback of interactivity, ease of use, usefulness, and motivation, we can see that the system can build a virtual simulation teaching scenario with good interactivity, which enables the test subjects to use the virtual experimental equipment and instruments easily, freely, and fluently to learn and explore sports knowledge independently and improve their sports knowledge, practical application ability, and learning motivation.

In this paper, two decision tree algorithms, ID3 and C4.5, are used to select the same training dataset to construct models of student reporting situations, and then the same test dataset is used to evaluate the performance of these two different models. The best performing algorithm is shown in Figure 10.

The accuracy of the two algorithms was calculated according to the formula for the accuracy assessment index, and Figure 10 shows that the accuracy of C4.5 is significantly higher than that of the ID3 algorithm.

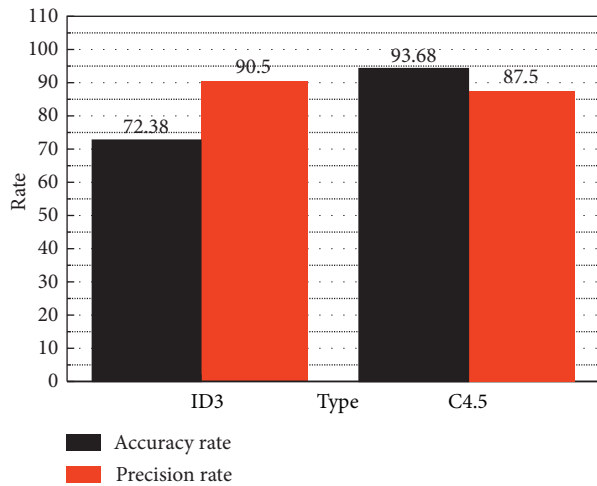


FIGURE 10: Accuracy rate and precision rate.

#### 4. Conclusion

In this paper, the CART algorithm is applied to the analysis and prediction of students' basic information, course test scores, and e-learning platform learning behavior characteristics. In addition, a CART model is constructed based on the improved decision tree algorithm to study the effectiveness of physical education in universities and to study its evaluation model and computer simulation. The CART identifies rules for correlating student learning factors, and some pedagogical strategies can be adjusted, such as the need to have a relatively effective teaching method in the e-learning platform, whether to make it mandatory or set open-ended assignments to guide students' independent learning. In addition to guiding the importance of foreseeing the building of the profession in terms of student and professional teaching, some degree of assessment of the effectiveness of student learning may be achieved. There are many different methods of data mining, among which decision trees and random forests have a very wide range of applications and practical implications. This is because they are theoretically clear, easy to understand, relatively computationally modest, and highly accurate. In this paper, the application of CART algorithms to the analysis and prediction of the assessment of student effectiveness in physical education is significant.

#### Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

#### Conflicts of Interest

The authors declare that they have no known financial conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

#### References

- [1] N. Segev, M. Harel, S. Mannor, K. Crammer, and R. El-Yaniv, "Learn on source, refine on target: a model transfer learning framework with random forests," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 9, pp. 1811–1824, 2017.
- [2] P. A. Vagefi, D. Bertsimas, R. Hirose, and N. Trichakis, "The rise and fall of the model for end-stage liver disease score and the need for an optimized machine learning approach for liver allocation," *Current Opinion in Organ Transplantation*, vol. 25, no. 2, pp. 122–125, 2020.
- [3] M. R. Nachiappan, V. Sugumaran, and M. Elangovan, "Performance of logistic model tree classifier using statistical features for fault diagnosis of single point cutting tool," *Indian Journal of Science and Technology*, vol. 9, no. 47, pp. 1–8, 2016.
- [4] S. M. M. Rubiano and J. A. D. Garcia, "Analysis of data mining techniques for constructing a predictive model for academic performance," *IEEE Latin America Transactions*, vol. 14, no. 6, pp. 2783–2788, 2016.
- [5] X. Ke, L. Qi, and C. Zeng, "A partitioned and asynchronous cellular automata model for urban growth simulation," *International Journal of Geographical Information Science*, vol. 30, no. 4, pp. 637–659, 2016.
- [6] S. Figini, R. Savona, and M. Vezzoli, "Corporate default prediction model averaging: a normative linear pooling approach," *Intelligent Systems in Accounting, Finance and Management*, vol. 23, no. 1-2, pp. 6–20, 2016.
- [7] S.-W. Lin, Y.-T. Liu, and M. A. Jerusalem, "Bayesian reliability analysis of a products of probabilities model for parallel systems with dependent components," *International Journal of Production Research*, vol. 56, no. 4, pp. 1521–1532, 2018.
- [8] P.-F. Pai, L.-C. Chen, and K.-P. Lin, "A hybrid data mining model in analyzing corporate social responsibility," *Neural Computing and Applications*, vol. 27, no. 3, pp. 749–760, 2016.
- [9] S. A. Mokeddem, "A fuzzy classification model for myocardial infarction risk assessment," *Applied Intelligence*, vol. 48, no. 5, pp. 1233–1250, 2017.
- [10] J.-P. Martin and D. Germain, "Dendrogeomorphic reconstruction of snow avalanche regime and triggering weather conditions," *Progress in Physical Geography: Earth and Environment*, vol. 40, no. 4, pp. 527–548, 2016.
- [11] G.-T. Gwak, S.-H. Ahn, J.-h. Kim, Y.-s. Weon, and O.-Y. Kwon, "Prediction model for the risk of scapular winging in young women based on the decision tree," *Physical Therapy Korea*, vol. 27, no. 2, pp. 140–148, 2020.
- [12] H.-K. Kim, K.-H. Choi, S.-W. Lim, and H.-S. Rhee, "Development of prediction model for prevalence of metabolic syndrome using data mining: Korea national health and nutrition examination study," *Journal of Digital Convergence*, vol. 14, no. 2, pp. 325–332, 2016.
- [13] W. B. Dieter, J. P. Collins, and A. A. Guccione, "Predicting outcomes within an innovative post-acute rehabilitation model for older adults," *BMC Geriatrics*, vol. 19, no. 1, pp. 1–9, 2019.
- [14] Y. Mao, M. Zhang, P. Sun, and G. Wang, "Landslide susceptibility assessment using uncertain decision tree model in loess areas," *Environmental Earth Sciences*, vol. 76, no. 22, pp. 1–15, 2017.
- [15] G. Tao, H. Song, J. Liu, J. Zou, and Y. Chen, "A traffic accident morphology diagnostic model based on a rough set decision tree," *Transportation Planning and Technology*, vol. 39, no. 8, pp. 751–758, 2016.

- [16] M. S. Yorgun and R. B. Rood, "A decision tree algorithm for investigation of model biases related to dynamical cores and physical parameterizations," *Journal of Advances in Modeling Earth Systems*, vol. 8, no. 4, pp. 1769–1785, 2016.
- [17] Y. Tanoue and S. Yamashita, "Loss given default estimation: a two-stage model with classification tree-based boosting and support vector logistic regression," *Journal of Risk*, vol. 21, no. 4, pp. 19–37, 2019.
- [18] X. Wang, Z. Guo, W. Zhang et al., "Study on the decision tree model for risk factors of vascular cognitive impairment," *Chinese Journal of Behavioral Medicine and Brain Science*, vol. 26, no. 6, pp. 534–538, 2017.
- [19] J. Zhao, "Study on the application of classification tree model in building the risk model for nosocomial infection in critically ill patients of emergency department," *Journal of Postgraduate Medicine*, vol. 40, no. 9, pp. 791–794, 2017.
- [20] D. K. Nagthane and A. M. Rajurkar, "An improved diagnosis technique for breast cancer using LCFS and TreeHiCARE classifier model," *Sensor Review*, vol. 39, no. 1, pp. 107–120, 2019.
- [21] R. Prasetya, "Data mining application on weather prediction using classification tree, naïve bayes and  $K$ -nearest neighbor algorithm with model testing of supervised learning probabilistic brier score, confusion matrix and ROC," *Journal of Applied Communication and Information Technologies*, vol. 4, no. 2, pp. 25–33, 2020.
- [22] S. Jia and Y. Pang, "Teaching quality evaluation and scheme prediction model based on improved decision tree algorithm," *International Journal of Emerging Technologies in Learning (IJET)*, vol. 13, no. 10, pp. 146–157, 2018.

## Research Article

# Real-Time Visualization Optimization Management Simulation of Big Data Stream on Industrial Heritage Cloud Platform

Mengya Gao 

*School of Civil Engineering and Architecture, Wuhan University of Technology, Wuhan 430070, China*

Correspondence should be addressed to Mengya Gao; 245238@whut.edu.cn

Received 24 August 2020; Revised 10 October 2020; Accepted 18 October 2020; Published 28 October 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Mengya Gao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recently, the development and utilization of industrial heritage resources by using big data has gradually attracted attention. This paper proposes a real-time visualization optimization management simulation of an industrial heritage cloud platform, which realizes the high reliability and diversified storage and utilization of industrial big data by the cloud data distributed storage subsystem. The big data prediction model of the GRU neural network based on a spark distributed framework is constructed to realize the prediction of industrial genetic data. Finally, visualization technology can provide information supporting for industrial production by displaying effective information intuitively. The model's effectiveness and reliability are verified by simulation.

## 1. Introduction

China's industrial development level has reached a new stage due to the gradual improvement and rapid development of today's social and economic system [1]. In order to better adapt to the industrial development environment of modern society, it is imperative to do a good job of industrial heritage information analysis, which is also an important research topic of modern industrial enterprise development [2, 3]. The protection and management of industrial heritage data is of great significance for enterprises to better analyse the industrial development history and the evolution process of industrial development [4]. Therefore, according to the existing big data information technology application, it is necessary to do a good job in the management and integration of industrial heritage resources, so as to provide effective industrial history development content for the development of modern industrial enterprises and make the modern industrial development conform to the development of the social system. The basic demands are of great significance to guide the industrial development [5]. Especially, in the era of big data, people's workload of processing information is increasing [6]. Therefore, in order to

effectively save the time of information processing and improve the efficiency of information processing, more and more fields begin to use visual methods to process data, and the main emphasis is on the comprehensive processing of some large-scale information [7]. The advent of the big data era is both an opportunity and a challenge for information processing. Therefore, relevant staff should strengthen the development of visualization technology so as to provide technical support for actual data processing [8].

At present, many scholars at home and abroad have carried out certain research and exploration on the application of industrial big data at the level of technical tools. The literature [9] has sorted out the current situation of the use of big data in industrial enterprises in the world, elaborated and summarized the current use needs of industrial big data and the difficulties to be overcome, and pointed out the key direction of its application research in the future. The paper elaborates the importance of industrial big data for the development of intelligent manufacturing, as well as the current development status of industrial big data in this field, and puts forward the current lack of industrial data standards, data security problems, and future development suggestions. The literature [10] studies the application status

of big data technology in industrial production workshop and summarizes the importance of big data technology in this field. The literature [11] analyses the current use of big data technology in industrial and mining enterprises from various aspects and analyses the challenges faced by big data in this field and the development direction in the future. In recent years, as a representative of the LSTM algorithm, the time series data prediction and analysis method based on the deep neural network model has accumulated certain research results. ARIMA (autoregressive integrated moving average model) time series model and other neural network algorithms are used to mine and analyse a kind of time series data [12]. However, the ARIMA method has strict requirements for data conditions and needs to meet the stability of the target data and cannot process the nonlinear information in the data to a certain extent. In the literature, the algorithm of combining CNN (convolutional neural networks) and RNN (recurrent neural networks) is used to transform keystroke time data into keystroke vectors so as to complete the learning of personal keystroke vector sequence data [13, 14]. In the literature [15], an improved RNN model based on the time series decomposition backward propagation algorithm is proposed, and a RNN prediction model based on time series decomposition is established, which improves the accuracy of prediction [16]. Although the RNN algorithm has a certain memory ability for the input of the past time, it cannot solve the problem of gradient vanishing or gradient explosion [17]. The LSTM neural network algorithm is used to predict the bus time from the starting point to the target location [18]. A prediction model based on RNN for obtaining information is proposed, which can achieve high accuracy prediction for messages [19]. Although the LSTM algorithm overcomes the problems of RNN gradient vanishing and gradient explosion, its structure is too complex and the model parameters are too many, so the training time is doubled. Especially in the face of large-scale data set analysis and processing requirements, this method is difficult to meet the actual computing speed requirements [20, 21].

Therefore, in order to solve the practical application problems of the above industrial big data technology and the performance problems of data mining algorithms, this paper constructs a small private cloud platform, which is built on the current mainstream Hadoop distributed computing platform. At the same time, to solve the algorithm performance problems of cloud platform analysis, it uses the GRU (gated recurrent unit) recurrent neural network combined with a spark distributed computing engine to realize the prediction and analysis of industrial time series big data, and the effective information is displayed by data visualization method. The simulation shows the effectiveness of the visualization platform.

## 2. Structure and Design of Industrial Heritage Cloud Platform

The industrial heritage big data cloud platform constructed in this paper is an intelligent monitoring and analysis big

data cloud platform integrating real-time monitoring and intelligent prediction analysis. Under the background of big industrial data acquisition and high efficiency [22], it is of great significance to solve the problems of traditional industrial data acquisition and high-efficiency data acquisition which has reference significance.

*2.1. Architecture Design of Cloud Platform.* According to the order of data analysis, the big data technology system is divided into three levels: data integration, data processing, and knowledge visualization, and the data processing process is shown in Figure 1.

The data processing process of the industrial big data cloud platform constructed in this paper can be roughly divided into data collection stage, data persistence stage, and data information processing to the final visual decision-making management. Firstly, in the data acquisition stage, the data from multiple data sources are collected into the industrial big data platform constructed in this paper, and the data are stored in different storage modes according to the characteristics and actual needs of different data. Then, in the data analysis stage, through the relevant data analysis tool components, the data are analysed and mined and further store the results. The analysis results of the previous step are finally visualized for the final decision management.

On the basis of data processing process, combined with the overall requirements, the overall architecture of cloud platform is designed from the perspective of services required in each stage. The overall level service of the cloud platform is shown in Figure 2. Its industrial genetic cloud platform includes data layer, data integration and storage layer, computing layer, data service layer, and application layer.

In the service of industrial heritage cloud platform which is shown in Figure 2, its work can be described as follows: the sensor equipment transmits the monitoring data to the data processing platform through HTTP, TCP, MQTT, and other transmission methods and stores it in MySQL after analysis, providing support for real-time query. The massive historical data use hive storage warehouse to provide large-scale data support for data mining and analysis of the system. In order to improve the query and analysis speed of the system, the query engine based on memory computing is used in statistical query and data mining analysis of historical data. In the application layer, the output of the big data platform layer is used for chart display. Build a web server platform to provide a visual interface for data display and analysis. The data processing of terminal equipment includes real-time monitoring, statistical query analysis, data mining analysis, and data visualization services.

The cloud data distributed analysis subsystem constructed in this paper has the following advantages:

- (1) The efficiency of the algorithm: the prediction algorithm based on the GRU threshold recurrent neural network has less parameters and better stability, which is very suitable for the prediction of industrial time series data.

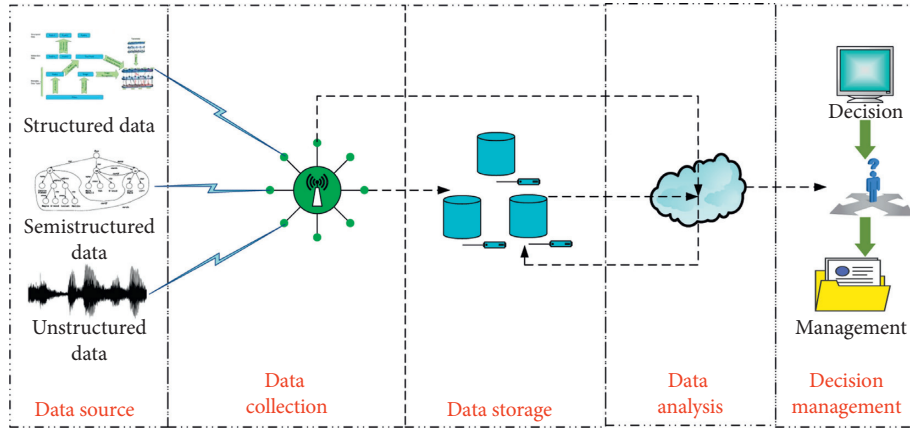


FIGURE 1: Process chart of data mining and visualization on the cloud platform.

- (2) High availability: the GRU threshold recurrent neural network is constructed on the basis of a spark core engine. It makes full use of spark's memory-based computing characteristics while parallel computing on multiple computers so that the algorithm still has high availability in the environment of large amount of data.
- (3) High scalability: the bottom layer of the cloud data distributed analysis subsystem is based on a spark distributed computing framework. In other words, the data analysis subsystem can be integrated with other components of the Hadoop ecosystem, which ensures the high scalability of the cloud platform in future application development.
- (4) Support of the cloud platform for other algorithms in the future: in terms of the data analysis algorithm, the cloud data distributed analysis subsystem itself is based on the design of a spark distributed computing platform, which makes it suitable for spark. At the same time, the subsystem supports the development and application of other types of data mining algorithms in the future

*2.2. Structure Design of Industrial Heritage Cloud Platform Based on Hadoop.* The structure of the industrial heritage cloud platform is shown in Figure 3. The platform is composed of a master node and a group of task subnodes. In the platform, the master node is responsible for the task scheduling and platform management and each subnode is responsible for the task execution, and after the task is completed, the data processing results are fed back to the master node, and the main node presents the results to the user and completes the serialization operation.

Firstly, considering the high reliability and high scalability of the industrial big data cloud platform, the cloud data distributed analysis subsystem is also designed as an independent middleware to provide data analysis services for the cloud platform. Secondly, the cloud data distributed analysis subsystem provides model training and data prediction and analysis functions for the industrial big data cloud platform as a whole. Therefore, the data analysis

subsystem provides two core functions: model training and cloud reserved model library. Finally, from the perspective of the overall performance of the industrial big data cloud platform, when facing the data analysis scenario with large amount of calculation and responsible calculation, the traditional single machine computing mode cannot meet the actual production requirements. Therefore, this part builds the data analysis subsystem on the spark platform and takes spark as the core computing engine. This design can make full use of the characteristics of spark parallel computing and memory-based computing advantages to improve the efficiency of data analysis.

### 3. Research on Big Data Mining Algorithm Based on GRU Network

*3.1. Data Mining Model Based on GRU.* In short-term industrial data forecasting, historical load series is the most important input. It contains rich information to forecast the future load demand and can represent the potential law of load demand. When the traditional machine learning method and single DNN are used to process historical load data, it is necessary to manually select relevant features from the original data, such as the load value of the previous hour and the load value of the same time of the previous day. The features selected according to the correlation will destroy the potential internal relations in the historical load series and affect the accuracy of the prediction results, and this feature selection method also increases the difficulty of prediction. The GRU neural network can effectively avoid this problem. Its internal gating cycle structure can automatically learn the relevant features from historical load data without manual extraction, which makes the prediction method more simple and easy to implement, and can also improve the prediction accuracy. In addition to the historical load series, short-term load forecasting is also affected by weather, holidays, dates, and other factors. The internal potential laws of these features are not obvious, so they are not suitable for the input of the GRU neural network. DNN can effectively deal with these external factors and learn the relationship between these characteristics and load demand so as to improve the prediction accuracy.

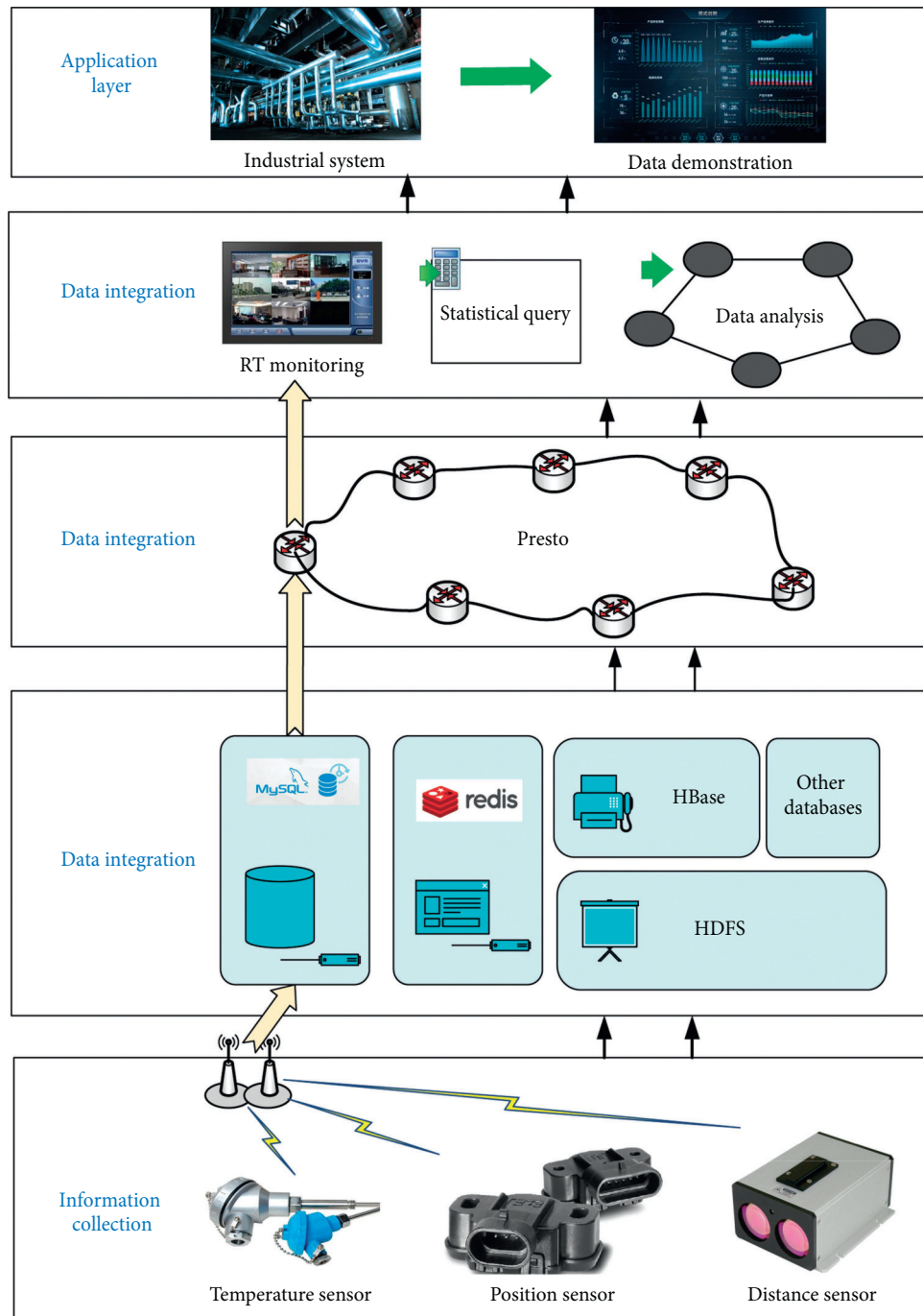


FIGURE 2: Architecture of the industrial heritage cloud platform.

The structure of the GRU prediction model is shown in Figure 4. The historical data input matrix  $T1$  in the model is the input of the GRU network, and the matrix  $TN$  is the output of the GRU network. The description of each layer in the input matrix model composed of other features is as follows.

The model is mainly divided into three layers. The first layer is data processing layer, which mainly normalizes the data and discretizes the normalized data into time series data; the second layer mainly extracts the features with small amount of data to reduce the prediction error; the third layer

mainly uses the prediction model based on GRU unit to send the data and feature extraction data to this layer, where the prediction of data can be realized. GRU unit is the core part of this model, and the detailed information is designed as follows. (1) Preprocess the historical input data and select the input features. According to the characteristics of historical data, external factors with greater influence are selected as other features. The length of historical load data and external characteristics is  $m$ , which needs to be corresponding to each other. In addition, in order to facilitate training, it is necessary to normalize the real value data and code the symbolic

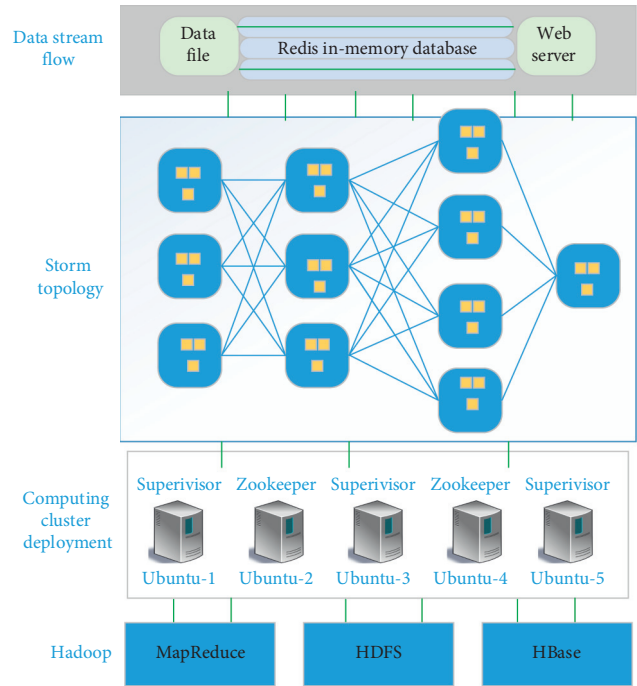


FIGURE 3: Architecture of the industrial heritage cloud platform based on Hadoop technology.

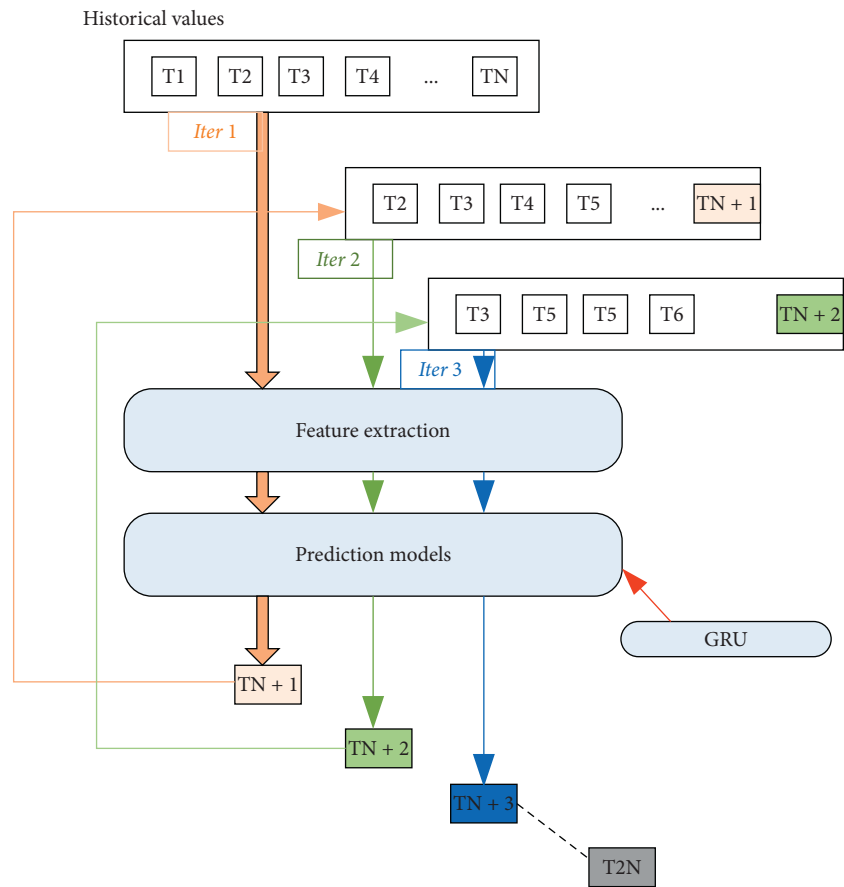


FIGURE 4: Architecture of the industrial heritage cloud platform based on Hadoop technology.



data. (2) The GRU model is trained by the training sample set, and the network parameters are optimized and updated to get the prediction model. The input data samples at the time to be predicted are input into the trained model to obtain the load forecasting value at each time. The output of a neuron in GRU networks is calculated as follows:

$$\begin{aligned} \mathbf{z}_t &= \sigma(\mathbf{W}^{(z)}\mathbf{x}_t + \mathbf{U}^{(z)}\mathbf{h}_{t-1}), \\ \mathbf{r}_t &= \sigma(\mathbf{W}^{(r)}\mathbf{x}_t + \mathbf{U}^{(r)}\mathbf{h}_{t-1}), \\ \tilde{\mathbf{h}}_t &= \tanh(\mathbf{W}\mathbf{x}_t + \mathbf{r}_t \circ \mathbf{U}\mathbf{h}_{t+1}), \\ \mathbf{o}_t &= \mathbf{h}_t = \mathbf{z}_t \circ \mathbf{h}_{t-1} + (1 - \mathbf{z}_t) \circ \tilde{\mathbf{h}}_t, \end{aligned} \quad (1)$$

where  $\mathbf{h}_{t-1}$  and  $\mathbf{h}_t$  are the output of hidden layer at time  $t - 1$  and  $t$ ;  $\mathbf{z}_t$ ,  $\mathbf{r}_t$ , and  $\mathbf{o}_t$  denote the output of update gate, reset gate, and neuron at time  $t$ ; and  $\mathbf{W}^{(z)}$ ,  $\mathbf{U}^{(z)}$ ,  $\mathbf{W}^{(r)}$ ,  $\mathbf{U}^{(r)}$ ,  $\mathbf{W}$ , and  $\mathbf{U}$  represent the weight parameters that we learned during the training process.

**3.2. Evaluation Index of Model.** In the training process, the model parameters are optimized by using the Adam (adaptive motion estimation) [21] optimization algorithm. Adam is a first-order optimization algorithm which can replace the traditional stochastic gradient descent process. In the process of training, the weight and deviation of each neuron node in the network model are updated iteratively to reduce the output value of loss function to the optimal value. The loss function of the model uses the mean square error function:

$$\min_{\theta=\Omega} \text{MAE} = \frac{1}{M} \sum_{i=1}^M |T_e^i(iT|\theta) - T_e^i(iT)|^2, \quad (2)$$

where  $T_e^i(iT|\theta)$  and  $T_e^i(iT)$  represent the predicted engine torque and real engine torque at sample time  $iT$ , respectively,  $i$  is the length of the  $i$ -th time step, and  $M$  is the total number of the sample time that are used to train or validate.  $\Omega$  is the solution space of  $\theta$ .

**3.3. Parallel Distributed GRU Prediction Model Based on Spark.** In order to speed up the training of the model, the spark distributed parallel computing framework is used in the training of the model. The idea of “divide and conquer” is adopted. The training data are distributed to a specified number of cluster worker nodes through the spark parallel computing framework. Each worker node executes its own according to the logic of the actual task. In the model training task, the driver reduces and averages the weight parameters of each node by setting the average frequency of parameters and redistributes the weight results to each node until the predetermined training target of the model is reached. Figure 5 shows the workflow of model parallelization training.

The detailed process is formulated as follows. Spark application communicates with spark cluster through Sparkccmtext in driver program. Spark program applies for the required resources from the cluster resource manager through Sparkccmtext. The cluster resource manager

allocates the resources and creates an exciter on each running node. Sprkcontext distributes the program code and corresponding data trained by the model to the running node, and each work node allocates the current allocation the model training task is parallelized. Finally, the results of each node are collected to the driver program to realize the parallel training of the model.

## 4. Simulation Results and Analysis

**4.1. Simulation Environment of Cloud Platform.** This part mainly introduces the data server and Hadoop cluster server node environment. The relevant environment and configuration information of the data server and cluster server nodes are shown in Table 1. All of our simulations are performed on the parallel cloud platform with one GTX-1080Ti card under CUDA9.0 and cudnnv7. In addition, the Sqoop component can realize the mutual conversion between the traditional relational database data and HDFS, HBase and other data, so as to ensure the efficient and safe import and export of data in different systems. The Spring-Boot OpcUa data acquisition server realizes the real-time data acquisition and monitoring management of remote equipment in the factory. Abp Core server is the basis of cloud platform visual management.

**4.2. Validation of GRU Evaluation Model.** The simulation takes a group as an example, which mainly provides OEM services to famous overseas clothing brand companies. The export revenue has occupied the first place of knitting apparel enterprises for many years and also maintains the first position of export to Japan for many years.

Planned production capacity, planned capacity index, service capacity index, timely delivery rate, order fulfilment rate, and average delivery period are the main indicators to evaluate a company. We use industrial heritage data for evaluation and visualization. In order to compare the advantages of our algorithm, we choose to compare it with SVM (support vector machine algorithm). The comparison results are shown in Figure 6.

As can be seen from Figure 6, external customers and internal management departments can provide reports intuitively and clearly on this unified interface so as to find out the changes of capacity indicators in time. The system will also rely on the interval early warning information to carry out automatic early warning prompt, meet the statistical data support of predictive scheduling, and monitor the manufacturing capacity under the contract. In addition, we can use the prediction model based on the GRU network to accurately predict the operation status of enterprises, and it is consistent with the actual production data. Compared with the SVM-based algorithm, our algorithm has higher prediction accuracy. The detailed prediction results are shown as in Figure 7.

From the results, it can be seen that the industrial heritage cloud platform model proposed in this paper has better prediction effect; it can more accurately capture the law of data change, can predict the trend of data change

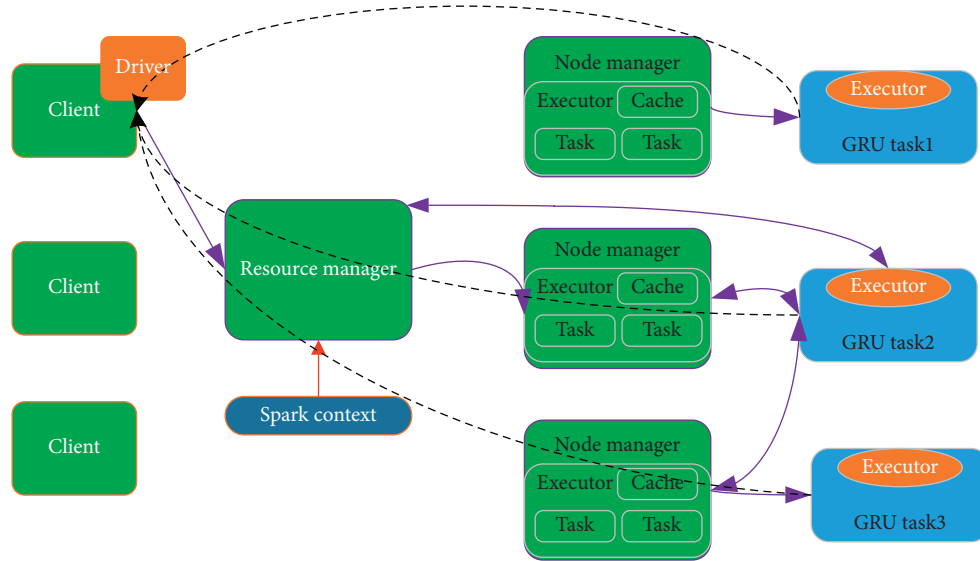


FIGURE 5: Parallel distributed GRU prediction model based on spark.

TABLE 1: Overall deployment of the cloud platform.

| Server type   | OS         | Memory (GB) | IP address   | Other information           |
|---------------|------------|-------------|--------------|-----------------------------|
| Data server 1 | Windows 7  | 4           | 10.33.46.45  | opc.tcp://10.33.46.45: 4401 |
| Data server 2 | Windows 7  | 4           | 10.33.46.96  | opc.tcp://10.33.46.45: 4401 |
| Master node   | CentOS 7.5 | 64          | 10.33.46.113 | Hadoop 2.7.3; Spark 1.5.1   |
| Slave node 1  | CentOS 7.5 | 64          | 10.33.46.114 | Hadoop 2.7.3; Spark 1.5.1   |
| Slave node 2  | CentOS 7.5 | 64          | 10.33.46.115 | Hadoop 2.7.3; Spark 1.5.1   |

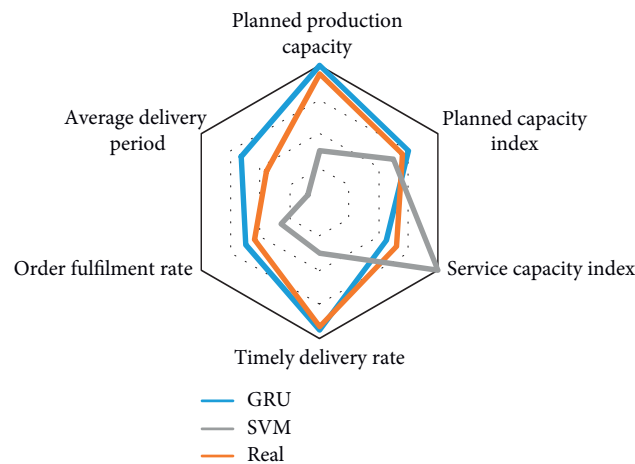


FIGURE 6: Simulation results compared with the SVM algorithm.

faster, can improve the prediction accuracy, and has high accuracy compared with the SVM prediction algorithm.

In addition, the cloud platform of industrial heritage also needs to evaluate the possible risks in the actual production process. This paper mainly shows the risk warning function of the industrial heritage cloud platform, and the results are shown in Figure 8.

Risk early warning was from the material and plan conflict, fabric supply interruption, lack of planning information, abnormal expectations, production progress lag,

and other five aspects of early warning. According to the contract number, the system can import production plan data and execution data according to the contract number, count the execution progress of all unfinished orders in real time, realize the automatic early warning of delivery date delay, and estimate the delayed delivery date under the existing constraints. As shown in the figure, our industrial heritage cloud platform can provide real-time early warning and visualization of possible problems in the production process so that enterprise decision makers can adjust their

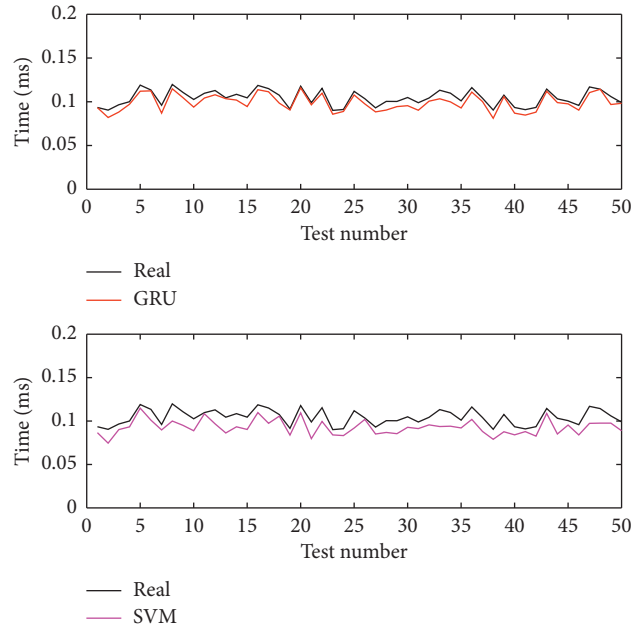


FIGURE 7: Simulation results compared with the SVM algorithm.

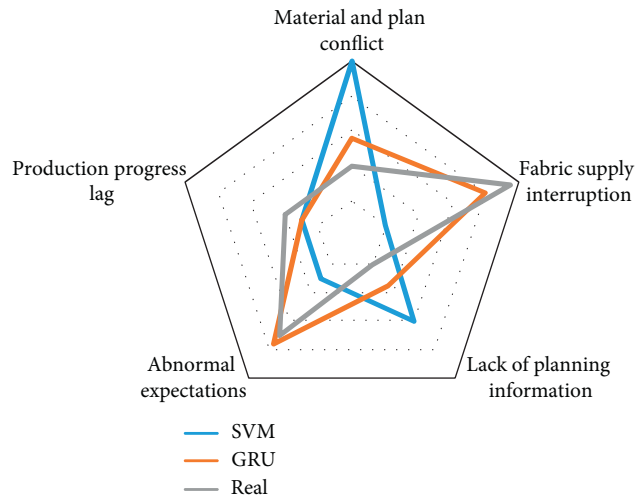


FIGURE 8: Enterprise operation risk prediction compared with SVM model.

own production plans. In addition, the accuracy of our prediction model is much higher than that based on SVM, which achieves good prediction performance.

#### 4.3. Production Data Visualization Based on Cloud Platform.

The process flow of products corresponding to different orders is not completely the same, and the operations required by different products in the same process are also different, so the decision and arrangement of production balance are complex. At the same time, the volatility of orders is also obvious. Orders fluctuate frequently due to the influence of seasons, fashion trends, prices, emergencies, and other factors. In addition, the company's OEM business is also affected by the order fluctuation. Different orders from

different enterprises have different order placing time and delivery time, and the product types and quantities presented in the weekly and daily production plans are also different. Considering the order satisfaction rate oriented production balance adjustment, the factors need to be considered comprehensively are more complex, which is a typical mixed loading balance decision-making problem under uncertain environment.

It can be seen from the figure that the production data are chaotic before processing, complex and disordered, as shown in Figure 9(a). We need to carry out feature extraction and visualization through the cloud platform, and the visualization results are shown in Figure 9(b). Through the full mining of the industrial heritage cloud platform, the relevance of production data began to show. Production data

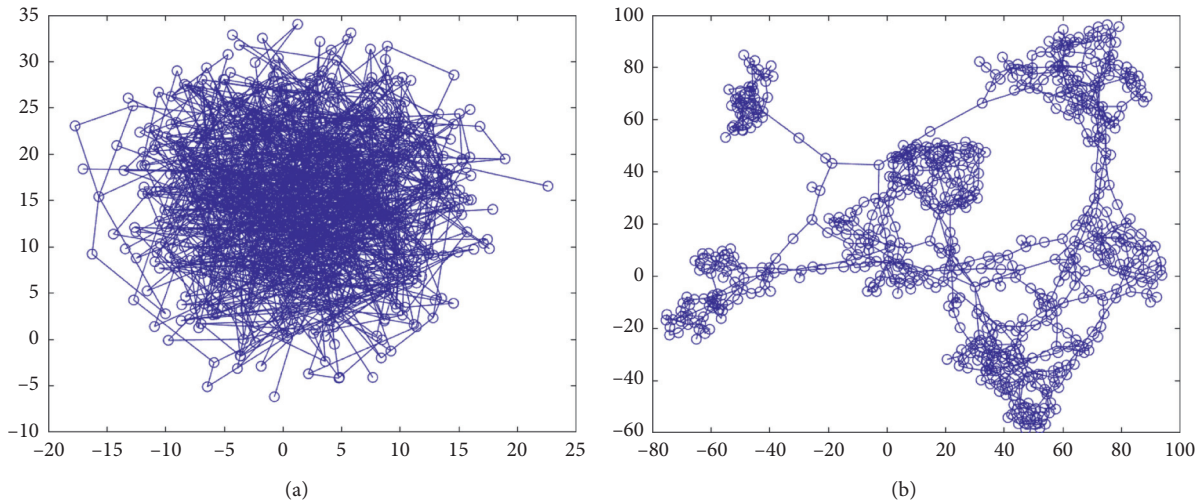


FIGURE 9: Production data visualization based on the cloud platform. (a) Before processing and (b) after processing.

are a complex high-dimensional and multi-dimensional data, which refer to data variables with multi-dimensional attributes. It widely exists in the application based on traditional relational database and data warehouse. The goal of high-dimensional multivariate data analysis is to explore the distribution rules and patterns of high-dimensional multivariate data items and to reveal the implicit relationship between different dimensional attributes. From the visualization results (Figure 9(b)), we can see that the production data mainly includes five aspects (planned production capacity, planned capacity index, service capacity index, timely delivery rate, order fulfilment rate, and average delivery period); we can figure out which leading force plays a leading role at what time and then deal with it according to the actual changes. Simulation results show that our industrial heritage big data platform has good data mining ability.

## 5. Conclusion

In order to solve the practical application problems of the above industrial big data technology and the performance problems of data mining algorithms, this paper constructs a small private cloud platform, which is built on the current mainstream Hadoop distributed computing platform. At the same time, to solve the algorithm performance problems of cloud platform analysis, it uses the gated recurrent unit recurrent neural network combined with a spark distributed computing engine to realize the prediction and analysis of industrial time series big data, and the effective information is displayed by data visualization method. The simulation shows the effectiveness of the visualization platform compared with the SVM algorithm.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no known conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

## References

- [1] K. Kaur, S. Garg, G. Kaddoum, E. Bou-Harb, and K. K. R. Choo, "A big data-enabled consolidated framework for energy efficient software defined data centers in IoT setups," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 4, pp. 2687–2697, 2020.
- [2] Y. Yan, C. Liu, X. Ma, and Y. Zhang, "Hardware-in-loop real-time simulation of electrical vehicle using multi-simulation platform based on data fusion approach," *International Journal of Distributed Sensor Networks*, vol. 15, no. 4, 2019.
- [3] T. Ranius, A. Hämäläinen, K. Eklöf et al., "The effects of logging residue extraction for energy on ecosystem services and biodiversity: a synthesis," *Journal of Environmental Management*, vol. 209, pp. 409–425, 2018.
- [4] S. Vallis, F. Galvez, M. Swidan et al., "Classical temples and industrial stores: survey analysis of historic unreinforced masonry (URM) precincts to inform urban seismic risk mitigation," *International Journal of Architectural Heritage*, vol. 12, no. 5–8, pp. 1276–1296, 2018.
- [5] P. Draus, S. Lovall, T. Formby, L. Baldwin, and W. Lowe-Anderson, "A green space vision in Southeast Michigan's most heavily industrialized area," *Urban Ecosystems*, vol. 22, no. 1, pp. 91–102, 2019.
- [6] F. Bruno, L. Barbieri, A. Lagudi et al., "Virtual dives into the underwater archaeological treasures of South Italy," *Virtual Reality*, vol. 22, no. 2, pp. 91–102, 2018.
- [7] M. J. García García, J. J. Escribano Otero, E. Millán Valdeperas, J. González Rodríguez, and F. Sánchez Carracedo, "Optimization of faculty time-management: some practical ideas," *The International Journal of Engineering Education*, vol. 34, no. 5, pp. 1467–1478, 2018.
- [8] W. J. Peluso, L. W. Zeng, C. K. Lin et al., "Application of an energy management system via the internet of things on a university campus," *Journal of Periodontology*, vol. 38, no. 3, pp. 1527–1534, 2017.

- [9] Z. Shen and V. Dinavahi, "Real-time MPSoC-based electro-thermal transient simulation of fault tolerant MMC topology," *IEEE Transactions on Power Delivery*, vol. 34, no. 1, pp. 260–270, 2019.
- [10] K. Dehghanpour and H. Nehrir, "Real-time multiobjective microgrid power management using distributed optimization in an agent-based bargaining framework," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6318–6327, 2018.
- [11] S. Samadzadehyazdi, Ansari, M. M. Mahdavi, and M. Bemaninan, "Significance of authenticity: learning from best practice of adaptive reuse in the industrial heritage of Iran," *International Journal of Architectural Heritage*, vol. 14, no. 3, pp. 329–344, 2020.
- [12] C. Arrighi, L. Rossi, E. Trasforini et al., "Quantification of flood risk mitigation benefits: a building-scale damage assessment through the RASOR platform," *Journal of Environmental Management*, vol. 207, pp. 92–104, 2018.
- [13] Y.-H. Lin, "The simulation of east-bound transoceanic voyages according to ocean-current sailing based on Particle Swarm Optimization in the weather routing system," *Marine Structures*, vol. 59, pp. 219–236, 2018.
- [14] S. Ciavarella, J.-Y. Joo, and S. Silvestri, "Managing contingencies in smart grids via the internet of things," *IEEE Transactions on Smart Grid*, vol. 7, no. 4, pp. 2134–2141, 2016.
- [15] A. A. Moses, K. Daisuke, and H. Sekyung, "Real-time demand side management algorithm using stochastic optimization," *Energies*, vol. 11, no. 5, pp. 1166–1173, 2018.
- [16] S. Sahoo, J. C.-H. Peng, A. Devakumar, S. Mishra, and T. Dragicevic, "On detection of false data in cooperative DC microgrids-A discordant element approach," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 8, pp. 6562–6571, 2020.
- [17] Y. Deng, W. Jiang, Z. Tang, Z. Ling, and Z. Wu, "Long-term changes of open-surface water bodies in the Yangtze river basin based on the Google Earth engine cloud platform," *Remote Sensing*, vol. 11, no. 19, p. 2213, 2019.
- [18] Y. Wang, Y. Lin, R. Y. Zhong et al., "IoT-enabled cloud-based additive manufacturing platform to support rapid product development," *International Journal of Production Research*, vol. 57, no. 11, pp. 3975–3991, 2019.
- [19] M. Kikuchi and K. Suzuki, "Characterizing vertical particle structure of precipitating cloud system from multiplatform measurements of a-train constellation," *Geophysical Research Letters*, vol. 46, no. 2, pp. 1040–1048, 2019.
- [20] H. Kong, Y. Fang, L. Fan, X. Zhang, H. Wang, and J. Hu, "A novel torque distribution strategy based on deep recurrent neural network for parallel hybrid electric vehicle," *IEEE Access*, vol. 7, no. 99, pp. 65174–65185, 2019.
- [21] F. Fei, Y. Zuo, L. D. Xu, L. Lv, and L. Zhang, "Internet of things and BOM-based life cycle assessment of energy-saving and emission-reduction of products," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 2, pp. 1252–1261, 2014.
- [22] B. P. Biswal, H. A. Vignolo-González, T. Banerjee et al., "Sustained solar H<sub>2</sub> evolution from a thiazolo [5, 4-d] thiazole-bridged covalent organic framework and nickel-thiolate cluster in water," *Journal of the American Chemical Society*, vol. 141, no. 28, pp. 11082–11092, 2019.

## Research Article

# Application of Empirical Orthogonal Function Interpolation to Reconstruct Hourly Fine Particulate Matter Concentration Data in Tianjin, China

Hongwu Zhou,<sup>1,2</sup> Haidong Pan,<sup>1,2</sup> Shuang Li ,<sup>3</sup> and Xianqing Lv <sup>1,2</sup>

<sup>1</sup>Physical Oceanography Laboratory, Qingdao Collaborative Innovation Center of Marine Science and Technology (CIMST), Ocean University of China, Qingdao, China

<sup>2</sup>Qingdao National Laboratory for Marine Science and Technology, Qingdao, China

<sup>3</sup>Ocean College, Zhejiang University, Zhoushan, China

Correspondence should be addressed to Shuang Li; [lshuang@zju.edu.cn](mailto:lshuang@zju.edu.cn) and Xianqing Lv; [xqinglv@ouc.edu.cn](mailto:xqinglv@ouc.edu.cn)

Received 28 April 2020; Revised 18 August 2020; Accepted 28 September 2020; Published 28 October 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Hongwu Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Fine particulate matter with diameters less than  $2.5\ \mu\text{m}$  (PM<sub>2.5</sub>) concentration monitoring is closely related to public health, outdoor activities, environmental protection, and other fields. However, the incomplete PM<sub>2.5</sub> observation records provided by ground-based PM<sub>2.5</sub> concentration monitoring stations pose a challenge to the study of PM<sub>2.5</sub> propagation and evolution model. Consequently, PM<sub>2.5</sub> concentration data imputation has been widely studied. Based on empirical orthogonal function (EOF), a new spatiotemporal interpolation method, EOF interpolation (EOFI) is introduced in this paper, and then, EOFI is applied to reconstruct the hourly PM<sub>2.5</sub> concentration records of two stations in the first half of the year. The main steps of EOFI here are to firstly decompose the spatiotemporal data matrix of the original observation site into mutually orthogonal temporal and spatial modes with EOF method. Secondly, the spatial mode of the missing data station is estimated by inverse distance weighting interpolation of the spatial mode of the observation sites. After that, the records of the missing data station can be reconstructed by multiplying the estimated spatial mode and the corresponding temporal mode. The optimal mode number for EOFI is determined by minimizing the root mean square error (RMSE) between reconstructed records and corresponding valid records. Finally, six evaluation indices (mean absolute error (MAE), RMSE, correlation coefficient (Corr), deviation rate bias, Nash–Sutcliffe efficiency (NSE), and index of agreement (IA)) are calculated. The results show that EOFI performs better than the other three interpolation methods, namely, inverse distance weight interpolation, thin plate spline, and surface spline interpolation. The EOFI has the advantages of less computation, less parameter selection, and ease of implementation, it is an alternative method when the number of observation stations is rare, and the proportion of missing value at some stations is large. Moreover, it can also be applied to other spatiotemporal variables interpolation and imputation.

## 1. Introduction

Fine particulate matter (PM<sub>2.5</sub>) is particulate matter with aerodynamic diameter less than  $2.5\ \mu\text{m}$  in ambient air [1]. Hazy weather will form if PM<sub>2.5</sub> concentration is too high, which has adverse impacts on human health, traffic, and outdoor activities [2], and it will also produce other indirect inestimable economic losses [3]. Therefore, many countries attach great importance to the monitoring and forecasting of PM<sub>2.5</sub> concentration. A large number of ground-based monitoring stations have been established. For example,

1500 monitoring stations have been set up in the United States. In China, around 1500 stations have been set up in 454 cities by 2018, and a new national ambient air quality standard for PM<sub>2.5</sub> was introduced in 2012 [1, 2]. Generally, it is believed that high PM<sub>2.5</sub> concentration has become a prominent challenge for air pollution control in China, which is mainly caused by the industrial combustion of coal and gasoline, traffic emissions, and long-distance transport [4, 5]. The North China Plain, especially the Beijing-Tianjin-Hebei region (Figure 1(a)), is one of the regions most severely affected by the hazy weather [4, 6]. To monitor air

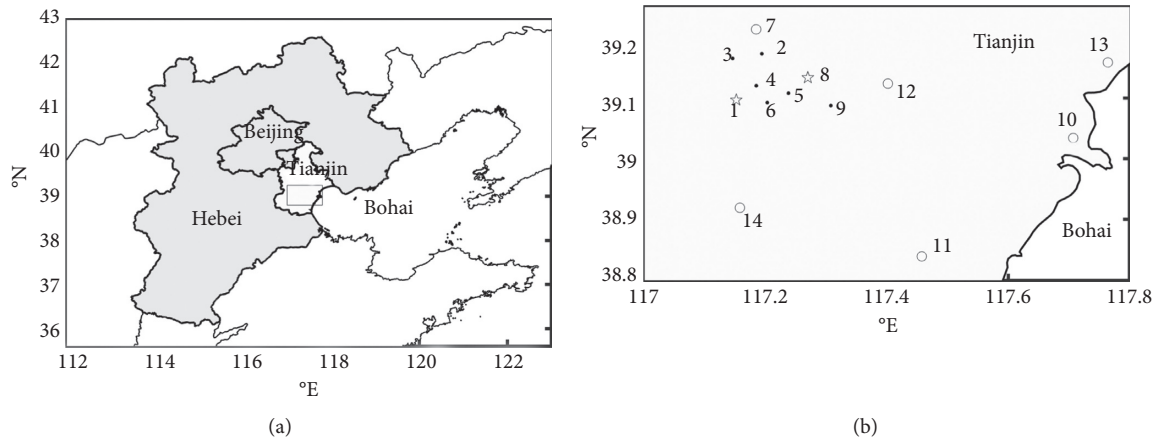


FIGURE 1: (a) Map of the Beijing-Tianjin-Hebei region. The rectangle in Tianjin is the study area of (b). (b) Location of 14 monitoring stations in Tianjin. From 1 to 14, they are located in the city testing center, Nankou Road, Qinjian Road, Nanjing Road, Dazhigu No. 8 Road, Qianjin Road, Beichen Science and Technology Park, Tianshan Road, Yuejin Road, Fourth Avenue, Yongming Road, Hangtian Road, Hanbei Road, and Tuanbowa. The stars represent the missing data stations (stations 1 and 8), the black dots represent the stations used for the interpolation (stations 2, 3, 4, 5, 6, and 9), and the circles represent the stations far from the missing data stations (stations 7, 10, 11, 12, 13, and 14).

pollution, many urban environmental stations have been built in this region, and many researchers have analyzed the causes and behavior of high PM<sub>2.5</sub> concentration recently [3, 7].

There have been many studies on PM<sub>2.5</sub> concentration data analysis methods, such as real-time data space interpolation of monitoring points, weighted regression models, and mixed models [1, 8]. The application of the preceding methods mostly depends on the complete and continuous monitoring data provided by local monitoring stations. However, problem arises when original spatiotemporal PM<sub>2.5</sub> concentration data are incomplete, which hinders further analysis and modelling, such as aerosol-related haze control and environmental health risk assessment [9, 10].

In practice, missing values and data gaps always exist in the original spatiotemporal observation records due to various factors. For example, satellite-based remote sensing may be affected by clouds, rain, aerosols, or incomplete track coverage in atmospheric research [11, 12]; *in situ* observations from land-based stations, shipborne monitoring, offshore buoy stations, and other platforms may suffer unexpected factors such as instrumental malfunction, power supply failure, and Internet outage [10, 13]. Directly ignoring incomplete spatiotemporal observation data should be carefully considered. The reasons include that the some platforms of data acquisition are expensive and irreplaceable (e.g., ocean research vessels and buoy stations), the demanding requirements of data quality (e.g., coastal tidal gauge records), and ignoring missing values sometimes may lead to biased spatial patterns and invalid inferences [10, 13]. Thus, many temporal, spatial, and spatiotemporal data interpolation and imputation methods have been proposed to fill these gaps in records.

Simple methods commonly used to fill gaps in univariate time series include mean value substitution (or median value and mode value), polynomial interpolation (linear, piecewise polynomials, and spline interpolations), and last

observation carried forward (*locf*), but they may result in large deviations when the time gaps are too large [14–17]. Based on a Markovian process, statistical parametric models include autoregressive (AR) models, moving average (MA) models, ARMA models, and linear weighted or exponential weighted MA. Complex machine learning techniques include gradient boosting and artificial neural networks (ANNs), which are computationally intensive [10, 18].

At present, there are also numerous spatial interpolation methods. Common simple methods include inverse distance weighting (IDW) interpolation [19], global polynomial interpolation (GPI), local polynomial interpolation (LPI) [20], surface spline (SS) interpolation [21], Cressman interpolation [22], and radial basis function (RBF). Using different basis functions, RBF includes thin plate spline (TPS), thin plate spline with tension, regularized spline, multiquadric spline, and inverse multiquadric spline. The TPS method does not need to set parameters, while other RBF needs to set parameters [23]. Some statistical-based methods (e.g., Kriging interpolation, optimal interpolation (OI), and Kalman filter) are conventional and classical methods in geoscience [12, 13, 24–27].

Numerous methods have been proposed to deal with spatiotemporal data containing missing values, and a considerable part of them are based on empirical orthogonal function (EOF) (e.g., [28–31]). Compared with other methods, EOF-based methods have the advantages of ease of implementation and less computation costs [32, 33].

EOF is based on the theory of matrix eigenvalue decomposition, and the core step of EOF is to decompose the spatiotemporal matrix into the sums of space-dependent spatial modes multiplied by corresponding time-dependent temporal modes. These EOF spatial and temporal modes can reveal data inherent characteristics or some phenomenon (e.g., ENSO) [13, 28]. EOF is usually used for spatiotemporal data analysis, but it can be also used to fill the missing data gaps.

One of the earliest applications of EOF interpolation is reconstruction of global-scale sea surface temperature (SST) [28]. Based on gridded data (1982–1993) processed by OI, EOF decomposition was performed to obtain spatial modes, and then, the temporal modes were expanded to longer time period (1950–1992) via least squares method when the data coverage was relatively poor; next, the longer time period spatiotemporal SST data were reconstructed. Their work can be considered as another form of optimal interpolation [13, 34]. In 2003, Data INterpolating Empirical Orthogonal Functions (DINEOF), an iterated EOF interpolation method, was proposed to fill the missing data gap [30]. Based on the principle of EOF, DINEOF was successfully used to reconstruct missing data and fill data gaps. Alvera-Azcárate et al. [32] reconstructed missing data of Adriatic sea surface temperature. Sirjacobs et al. [35] used DINEOF to show the reconstruction of complete space-time information for 4 years of surface chlorophyll-*a* (CHL), total suspended matter, and SST over the Southern North Sea and the English Channel. However, DINEOF may fail if the data gaps are too huge.

Similar to the principle of DINEOF, EOF interpolation (EOFI) was proposed to reconstruct spatially continuous water levels in the Columbia River Estuary using limited tide gauges along the river [36]. Their main steps are as follows: firstly, the spatial-temporal data matrix of the river existing observation stations was decomposed with EOF method. Then, Pan and Lv adopt one-dimensional linear interpolation and one-dimensional spline interpolation to estimate the missing data station's spatial modes, respectively; then, EOFI reconstruction sequence is obtained by the estimated spatial modes multiplied by corresponding temporal modes, and this reconstruction sequence was in good agreement with that of the NS\_TIDE method. NS\_TIDE is specially designed and applied to the analysis of river tidal water level, and river flow discharge data are needed [37].

Based on the research of Pan and Lv [36], this study attempts to extend the missing data station's EOFI spatial mode from one-dimensional spatial interpolation to two-dimensional spatial interpolation. The river upstream and downstream sites are nearly one-dimensional distributed, and there is a strong correlation between the upstream and downstream water level records (e.g., when the upstream of a river rises, the water level in the downstream generally rises). Therefore, it is reasonable to apply one-dimensional interpolation to establish the spatial mode's connection between the observation stations and the missing data station. Compared with the river water level reconstruction, the PM2.5 stations' correlation is not so strong and intuitive because the PM2.5 concentration stations are spatially distributed. To establish a connection between variables that two-dimensional distributed in space, a simple idea is using IDW, so EOFI here uses IDW to estimate the spatial modes of the missing data station. Of course, other spatial interpolation methods can also be applied to the establishment of spatial mode relationships, but we will not discuss them in this paper. We consider the simple case (IDW) to verify the usability of EOFI. To the best of our knowledge, our proposed EOFI has not been applied to PM2.5 concentration

data reconstruction currently; therefore, we firstly introduce and use this method to fill the data gaps and compare the result with IDW interpolation, surface spline (SS), and TPS interpolation. The competing methods we choose here are all widely used and easy to implement [38].

Compared with widely used DINEOF- and other EOF-based methods, the novelty of our method is to deal with the case of sparsely distributed observation stations and a large proportion of missing values in some stations' records. In this case, the data of the station with too many missing values are not suitable for EOF decomposition (DINEOF fills these gaps with first guess values and then uses these data for EOF decomposition); otherwise, the accuracy of temporal and spatial modes will be affected. EOFI here only uses the observation data with a small proportion of missing values for decomposition; thus, the EOF decomposed temporal and spatial modes are more accurate and less affected. Then, spatial interpolation is applied to establish spatial modes' connection between observation stations and missing data station, and next, the reconstruction sequence with optimal mode number is determined by root mean square error (RMSE). The EOFI reconstruction sequence can be used as a reasonable first guess value of the missing data station for other methods further EOF decomposition (e.g., DINEOF). In this way, the spatial mode patterns are considered to some extent. Further comparison between DINEOF and EOFI will be explained in Discussion.

The paper is arranged as follows: Section 2.1 describes the study area and data. Then, we revisit the principle of EOF decomposition and introduce IDW, EOFI, TPS, and SS. The evaluation indices of these methods will also be mentioned in Section 2. Four methods (IDW, EOFI, TPS, and SS) are applied to reconstruct two stations' PM2.5 concentrations records, and then, the results are compared with corresponding valid observations in Section 3. EOFI inverse distance weighting power  $P$ , the impact of site number and data time length on the EOFI reconstruction, and comparison between DINEOF and EOFI will be discussed and analyzed in Section 4. Finally, we present the advantages and disadvantages of EOFI in Section 5.

## 2. Materials and Methods

**2.1. Study Area and Data.** There are 14 monitoring stations (Figure 1(b)) located in Tianjin. These stations are distributed in different regions of the city: some stations are located in the urban area (e.g., stations 1, 2, and 3), while other stations are near the Bohai Sea (e.g., stations 10, 11, and 13). The PM2.5 concentration data provided by these monitoring stations come from China National Environmental Monitoring Center (CNEMC). The CNEMC releases near real-time PM2.5 concentration data online, but there are no direct data download interface [10]. Bai et al. used web crawler technology to obtain many cities PM2.5 concentration data from 2014 to 2019. Here, our data sources and acquisition method are the same. In this study, some of the stations provided the hourly PM2.5 data throughout the year of 2015, except for the first 25 hours from January 1st 0:00 AM to January 2nd 0:00 AM. Thus, the total time length is



8735 hours (8760 hours in 2015). The reason for first 25 hours missing values may be web crawler technology failure, or CNEMC did not release the data for that time period. Figure 2 shows the original observation records of several stations used in this study. Among them, the first half year PM2.5 concentration data of station (sta) 1 and station (sta) 8 are reconstructed and compared with their corresponding valid records (Figure 2 (1 and 7)). There are no observed data from June 30th 23:00 PM to the end of the year (near six months) in sta 1 and sta 8. In addition, Bai et al. [10] mentioned that some monitoring stations across China have stopped releasing PM2.5 observations since the middle of 2015, and consequently, observations at these stations for the second half of 2015 are missing. This is the exact case at sta 1 and sta 8 in Tianjin. In sta 1, 10.70% of the data in the first half year are missing, and the percentage of missing data for the nearly whole year record is 55.86% (Figure 2 (1)). At sta 8, the proportions of missing data for the first half year and the nearly whole year are 9.59% and 55.31%, respectively (Figure 2 (7)). It shows that there are still nearly 400 missing values in the first half of the year for both sta 1 and sta 8.

## 2.2. Methods

**2.2.1. EOF Decomposition.** The EOF method was firstly proposed by the statistician Pearson in 1902, and meteorologist Lorenz firstly introduced the EOF method into meteorological and climatic research in 1956 [39]. We consider that there are  $N$  stations providing observation records with data length  $L$ , composing the  $N \times L$  space-time matrix  $\mathbf{X}$ . The column  $\mathbf{x}_i$  consists of  $N$  points records at time  $i$  ( $i = 1, 2, \dots, L$ ). The most important step of EOF is to solve the eigenvalues and eigenvectors of symmetric matrix  $\mathbf{X}\mathbf{X}^T$ ; the results of this decomposition include eigenvalues  $\lambda_k$  and their corresponding eigenvectors  $\mathbf{F}_k$  (normalized orthogonal spatial modes) [13]:

$$\mathbf{X}\mathbf{X}^T\mathbf{F}_k = \lambda_k\mathbf{F}_k, \quad k = 1, \dots, N. \quad (1)$$

The column  $\mathbf{F}_k$  of matrix  $\mathbf{F}$  is arranged from left to right in the descending order of the corresponding eigenvalues  $\lambda_k$  ( $k = 1, \dots, N$ ), the elements of the diagonal matrix  $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$  are also arranged in this order, and thus, equation (1) can be written as follows:

$$\mathbf{X}\mathbf{X}^T\mathbf{F} = \mathbf{F}\mathbf{D}. \quad (2)$$

The  $N \times N$  matrix  $\mathbf{F}$  is called spatial modes coefficient matrix, which is also orthogonal (i.e.,  $\mathbf{F}\mathbf{F}^T = \mathbf{F}^T\mathbf{F} = \mathbf{I}$ ), corresponding to the temporal modes coefficient matrix  $\mathbf{A}$  or principal component (PC). The  $N \times L$  matrix  $\mathbf{A}$  is calculated by the following equation:

$$\mathbf{A} = \mathbf{F}^T\mathbf{X}. \quad (3)$$

The column vector  $\mathbf{x}_i$ ,  $N$  points records at time  $i$ , is reconstructed as

$$\mathbf{x}_i = \mathbf{F}\mathbf{a}_i. \quad (4)$$

Here,  $\mathbf{a}_i$  is the column of  $\mathbf{A}$  at time  $i$ , and obviously,  $\mathbf{X} = \mathbf{F}\mathbf{A}$ . The  $k$ -th row of the matrix  $\mathbf{A}$  is called the temporal  $k$ -th mode, and the element of the  $i$ -th column is the temporal coefficient at time  $i$ . Correspondingly, the column  $\mathbf{F}_k$  is called the spatial  $k$ -th mode, and the elements of the  $j$ -th row of  $\mathbf{F}$  (i.e.,  $\mathbf{F}(j)$ ) represent the coefficients of each spatial mode of the  $j$ -th station. Thus, matrix element  $\mathbf{F}_{jk}$  is the  $k$ -th spatial mode of the  $j$ -th station. The temporal modes are time-dependent, while spatial modes are space-dependent [13]. In addition, different spatial modes and different temporal modes are, respectively, orthogonal (i.e.,  $\mathbf{F}\mathbf{F}^T = \mathbf{F}^T\mathbf{F} = \mathbf{I}$  and  $\mathbf{A}\mathbf{A}^T = \mathbf{D}$ ). Finally, the eigenvalue  $\lambda_j$  of the  $j$ -th mode can be used to calculate the cumulative variance contribution rate of the first  $k$  modes to the total variance:

$$G(k) = \frac{\sum_{j=1}^k \lambda_j}{\sum_{j=1}^N \lambda_j} \times 100\%, \quad (k \leq N). \quad (5)$$

The closer the  $G(k)$  approaches 100%, the more information the first  $k$  modes reflect of the original signals [36]. In spatiotemporal data analysis, we often only care about the first  $k$  modes with large variance contribution and regard them as the dominant modes. However, many EOF-based interpolation methods do not only consider the dominant modes, and the less important modes should also be considered. The optimal number of modes for reconstruction is determined by the root mean square error between the reconstruction sequence and the corresponding valid observation record [40].

**2.2.2. IDW and EOFI.** The IDW formula is given as follows:

$$W_j = \frac{1/d_j^P}{\sum_{j=1}^N 1/d_j^P} \quad (j = 1, \dots, N), \quad (6)$$

$$\tilde{Z}_{\text{IDW}} = \sum_{j=1}^N Z_j \cdot W_j, \quad (7)$$

$$\tilde{X}_{\text{IDW}} = \sum_{j=1}^N \mathbf{X}(j) \cdot W_j, \quad (8)$$

where  $d_j$  denotes the distance between the  $j$ -th station and the target station,  $P$  is the inverse distance power parameter,  $W_j$  is the corresponding normalized weight,  $\mathbf{X}(j)$  denotes the observation records sequence at the  $j$ -th station (i.e., the  $j$ -th row of  $\mathbf{X}$ ), and  $\tilde{Z}_{\text{IDW}}$  and  $\tilde{X}_{\text{IDW}}$  represent IDW estimated value and estimated reconstruction sequence, respectively. IDW is based on Tobler's First Law of Geography: "everything is related to everything else, but near things are more related than distant things" [41]. The feature of this method is to produce "bull's eyes" around the observation points in the nearby area when observation points are rare and distributed sparsely [20]. For IDW, the common values of  $P$  are 1 and 2 (also called inverse squared distance weighting), so we only discuss the influence of these two parameters on IDW and EOFI in the later experiments.

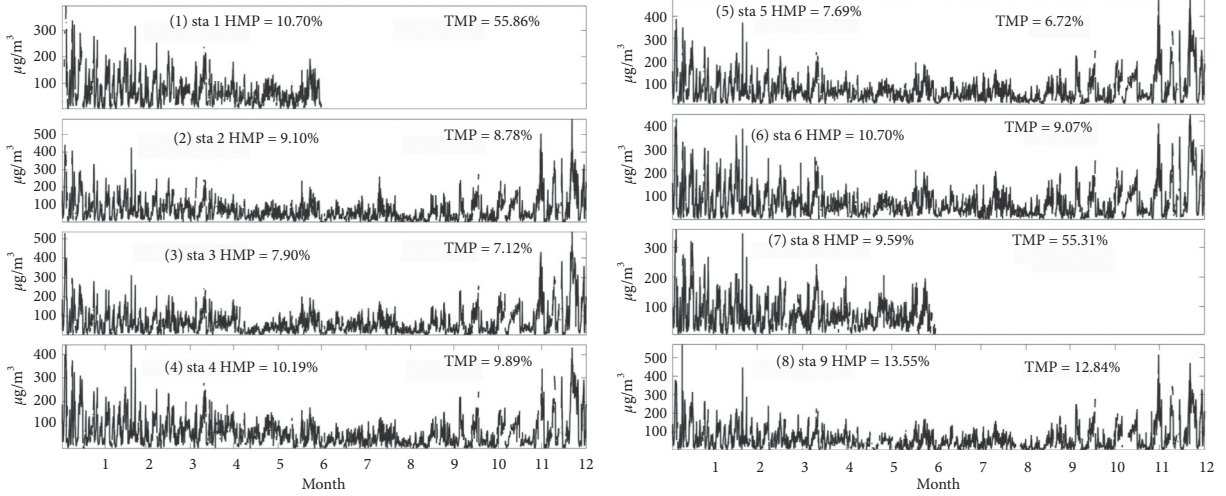


FIGURE 2: The original 12-month-long PM2.5 data records of 8 stations (sta 1, 2, 3, 4, 5, 6, 8, and 9) used in this study. The first half year missing percentage (HMP), i.e., the percentage of missing data in the first 4318 hours, and the total missing percentage (TMP), i.e., the percentage of missing data in 8735 hours, of each station are marked on the subgraphs.

In this study, the EOFI method steps are as follows: the missing data station shares the same temporal modes with observation stations, but the spatial modes  $\tilde{F}$  are estimated by the IDW interpolation of spatial modes of observation stations ( $\mathbf{F}(j)$ ,  $j = 1, \dots, N$ ):

$$\tilde{F} = \sum_{j=1}^N \mathbf{F}(j) \cdot W_j. \quad (9)$$

Here,  $W_j$  is the same as the weight mentioned in IDW (equation (6)). Then, the  $1 \times N$  row vector  $\tilde{F}$  and corresponding temporal mode  $\mathbf{A}$  reconstruct the estimated value  $\mathbf{x}_i^k$  at time  $i$  and estimated reconstruction sequence  $\tilde{\mathbf{X}}_{\text{EOF}}^k$  using the first  $k$  modes:

$$\begin{aligned} \mathbf{x}_i^k &= \tilde{F}(1, 1:k) \mathbf{A}(1:k, i), \quad 1 \leq k \leq N, \\ \tilde{\mathbf{X}}_{\text{EOF}}^k &= \tilde{F}(1, 1:k) \mathbf{A}(1:k, 1:L), \quad 1 \leq k \leq N. \end{aligned} \quad (10)$$

Using first  $k$  modes means that only the first  $k$  columns of  $\tilde{F}$  and the first  $k$  rows of  $\mathbf{A}$  are considered. Finally, the optimal mode number for EOFI reconstruction is determined by the minimizing RMSE between the reconstructed sequence ( $\tilde{\mathbf{X}}_{\text{EOF}}^k$ ,  $k = 1, \dots, N$ ) and the corresponding valid observation sequence  $\mathbf{X}_{\text{vid}}$ :

$$\tilde{\mathbf{X}}_{\text{EOF}} = \min_{\text{RMSE}(\mathbf{X}_{\text{vid}, \cdot})} \left\{ \tilde{\mathbf{X}}_{\text{EOF}}^k, k = 1, \dots, N \right\}. \quad (11)$$

The spatial mode is deemed space-dependent and can reflect the spatial characteristics under the assumption of EOF decomposition. In this study, the estimated spatial mode  $\tilde{F}$  is closely related to the distance from the observation station. If the missing data station and the observation station are close in space, their spatial modes are also close to each other (larger weight, equation (6)); thus, the EOFI reconstruction sequence is also close to the observation sequence, which is consistent with our experience.

Prior to reconstruction, the raw data matrix  $\mathbf{X}$  may contain missing values and cannot be directly EOF decomposed.

Therefore, it is necessary to preprocess the raw data and get the data matrix without missing measured value before decomposition. Here, we first replace the missing values with observed values' space average at missing values time points and then apply linear interpolation to fill all the temporal intervals (i.e., spatial mean value substitution and temporal linear interpolation). Note that the temporal gaps should not be too large, so as to avoid that the interpolation affects the accuracy of dominant temporal and spatial modes [36]. In this study, the data used for EOF decomposition include the preprocessed records of stations 2, 3, 4, 5, 6, and 9 (near one year). Their temporal gaps of original records are short (Figure 2 (2–6, 8)), so we believe that the dominant modes are slightly affected and still reliable. The first half year records of sta 1 and sta 8 are both excluded from EOF decomposition.

**2.2.3. Thin Plate Spline Method and Surface Spline.** The TPS method is a spatial interpolation method based on surface fitting, and it is one of the most frequently compared spatial interpolation methods [38], which was first proposed by Duchon [42]. It is often used to deal with uneven data in geoscience, such as generating continuous smooth elevation surface from discrete and sparse sample point elevation data. By simulating the bending of sheet metal, the TPS method generates a smooth surface with minimum bending energy through all observation points. Its form is as follows:

$$\tilde{Z}_{\text{TPS}} = \sum_{i=1}^N T_i d_i^2 \ln(d_i) + a + bx + cy. \quad (12)$$

Among them,  $d^2 \log(d)$  term is the basic function and  $a + bx + cy$  is the local trend function. The missing data station's horizontal coordinate  $(x, y)$  and its distances from the  $i$ -th ( $i = 1, \dots, N$ ) observation station are needed for TPS. In order to determine the  $N + 3$  unknown parameter  $T_i$  ( $i = 1, \dots, N$ ),  $a$ ,  $b$ , and  $c$  (equation (12)) are subject to the following relations:

$$\begin{aligned}
& \sum_{j=1, j \neq i}^N T_j d_{ji}^2 \ln(d_{ji}) + a + bx_i + cy_i = Z_i, \\
& \sum_{i=1}^N T_i = 0, \\
& \sum_{i=1}^N T_i x_i \\
& \sum_{i=1}^N T_i y_i \\
& d_{ji}^2 = (x_j - x_i)^2 + (y_j - y_i)^2,
\end{aligned} \tag{13}$$

with the  $N$  observation points' horizontal coordinates ( $x_i, y_i, i = 1, \dots, N$ ), distances between each other ( $d_{ij}, i, j = 1, \dots, N$ ), and observation values ( $Z_i, i = 1, \dots, N$ ), a smooth surface ( $N+3$  linear equations and  $N+3$  unknown parameters) is generated, the value at the missing data station is also assumed to be on this surface, and then, the TPS estimated value  $\tilde{Z}_{\text{TPS}}$  is calculated by equation (12). The TPS matrix form was fully described in Bookstein [43], and the coefficient matrix of unknown parameter is only related to spatial attributes (coordinate and distance), but not to time attribute.

The surface spline (SS) method is also a good spatial interpolation method based on surface fitting. It generates smooth surfaces through discrete points too. However, the basic function of the SS method is different from TPS. It does not consider trend term, the fitting function is different, and the radius  $R$  is introduced. Guo et al. [44] used the SS method to interpolate the bottom friction coefficient of the selected independent points to obtain values for the entire Bohai Sea and combined the adjoint assimilation method to invert the bottom friction coefficient of the entire sea. The SS method is also used for the inversion of initial conditions and parameters estimation in the ocean pollutant transport model [21], which is a significant improvement over the Cressman interpolation. Its form is as follows:

$$\tilde{Z}_{\text{SS}} = \sum_{j=1}^N S_j \left( \frac{d_j^2}{R^2} \ln \frac{d_j^2}{R^2} + 1 - \frac{d_j^2}{R^2} \right), \tag{14}$$

$$d_j^2 = (x - x_j)^2 + (y - y_j)^2. \tag{15}$$

Similar to TPS, the  $N$  observation points' spatial attributes and observation values sequences  $\mathbf{z}$  generate a smooth surface, and then, the unknown parameter column vector  $\mathbf{s}$  is solved by the matrix form:

$$\begin{aligned}
\mathbf{D}\mathbf{s} &= \mathbf{z}, \\
\mathbf{D} &= (\mathbf{D}_{ij})_{N \times N}, \\
\mathbf{s} &= (S_1, \dots, S_N)^T, \\
\mathbf{z} &= (Z_1, \dots, Z_N)^T, \\
\mathbf{D}_{ij} &= \begin{cases} \frac{d_{ij}^2}{R^2} \ln \frac{d_{ij}^2}{R^2} + 1 - \frac{d_{ij}^2}{R^2}, & i \neq j, \\ 1, & i = j. \end{cases}
\end{aligned} \tag{16}$$

Here, the elements of parameter matrix  $\mathbf{D}$  are only related to the distance between observation points  $d_{ij}$  ( $i, j = 1, \dots, N$ ) and prescribed radius  $R$ . The radius  $R$  is set to 15 km because the distance between any two stations is within this radius. After solving the unknown sequence  $\mathbf{s}$ , SS estimated value  $\tilde{Z}_{\text{SS}}$  of missing data station is calculated with equations (14) and (15). Note that the value of  $\mathbf{s}$  changes with radius  $R$ , but selecting  $R$  within the appropriate range will not have a great impact on the final interpolation result.

**2.3. Evaluation Indices.** At the end of Section 2.2.2, the preprocessing of the original data has been mentioned. We emphasize that the preprocessed data used for each interpolation method is the same. Therefore, the evaluation of different interpolation methods is persuasive and reliable. Table 1 summarizes their parameter settings. We will list a series of quantitative indices to evaluate these interpolation methods [38]. The evaluation indices listed in this study include mean absolute error (MAE), root mean square error (RMSE), correlation coefficient (Corr), and deviation rate bias, Nash–Sutcliffe efficiency (NSE) [45], and index of agreement (IA) (or Willmott's D) [46].

Among them, MAE (equation (17)) and RMSE (equation (18)) are often used as indicators of the performance of interpolation or models [38]. The smaller they are, the better the interpolation effect is. Corr (equation (19)) and bias (equation (20)) measure the correlation and deviation between simulation value sequence  $S$  and the observation series  $O$ , and  $\bar{S}$  and  $\bar{O}$  are their average values, respectively. Higher degree of correlation and smaller deviation both indicate the better interpolation effect. NSE (equation (21)) is a common index used to measure the performance or interpolation effect in meteorological, hydrological, and environmental models. Its value ranges from negative infinity to 1. The closer to 1, the simulation results are closer to observations; the closer to 0, the result are closer to the observation average values, but the process error is large, while negative NSE indicates that the performance of mean observed values is even better than simulated values and indicates this simulation unacceptable. IA (equation (22)) is referred as the potential error. IA is a nondimensional and bounded index with values closer to 1 indicating better agreement. The above six indices are defined as follows:

$$\text{MAE} = \frac{\sum |S - O|}{n}, \tag{17}$$

$$\text{RMSE} = \sqrt{\frac{\sum (S - O)^2}{n}}, \tag{18}$$

$$\text{Corr} = \frac{\sum (S - \bar{S})(O - \bar{O})}{\sqrt{\sum (S - \bar{S})^2} \sqrt{\sum (O - \bar{O})^2}}, \tag{19}$$

$$\text{bias} = \frac{\sum |S - O|}{\sum O} \times 100\%, \tag{20}$$

TABLE 1: Four interpolation methods parameter setting.

| Method | Parameter setting                |
|--------|----------------------------------|
| IDW    | $P=1$ and $2$                    |
| EOF1   | Spatial mode is dependent on IDW |
| TPS    | —                                |
| SS     | $R=15$ km                        |

$$\text{NSE} = 1 - \frac{\sum (S - O)^2}{\sum (O - \bar{O})^2}, \quad (21)$$

$$\text{IA} = 1 - \frac{\sum (S - O)^2}{\sum (|O - \bar{O}| + |S - \bar{O}|)^2}. \quad (22)$$

In Section 3.2, we calculated the above six evaluation indicators, which reflect the accuracy of these simulations, and the indicators for the EOF1 first  $k$  modes ( $k=1, \dots, N$ ) are also calculated. The results of EOF1 with the optimal mode number will be compared with other three interpolation methods.

**2.4. Site Selection.** To pursue better interpolation performance, here we just choose the data of the five nearest stations for interpolation; that is, the imputation of sta 1 and sta 8 data is based on the data of stations 2, 3, 4, 5, and 6 and the data of stations 2, 4, 5, 6, and 9 (Figure 1(b)), respectively, while the data of other stations are not included. The near one-year records of sta 1 and sta 8 are reconstructed, respectively, by interpolating data of the five nearest stations with four interpolation methods, and then, the reconstructed sequences are compared with corresponding valid observation data in the first half year (Figure 2) to calculate the evaluation index. In Section 4.2 for further validation, multiple sets of experiments in different time periods are implemented, and the RMSE between four interpolation methods' reconstruction sequence and corresponding valid observation records are further compared.

### 3. Results

**3.1. Interpolation Result of Four Methods.** The distances between the observation stations and the target station and the corresponding normalized weight are presented in Table 2. The distance from sta 4 is the shortest, and the weight is the largest in the sta 1 group, while distance from sta 5 is the shortest, and the weight is the largest in the sta 8 group. With the increase in IDW and EOF1 power parameter  $P$  (from 1 to 2), the normalized weights of the nearest stations (sta 4 and sta 5) increase, while the weights of other stations decrease. Therefore, the estimated spatial mode  $\bar{F}$  of sta 1 and sta 8 calculated by equation (9) is more affected by those of sta 4 and sta 5, respectively.

The temporal modes or principal components (PCs) of sta 1 and sta 8 (Figure 3) and the corresponding spatial modes (Table 2) are obtained by EOF decomposition. It can be seen that the variance contribution rate of PC1 in sta 1 and sta 8 is both over 98%, and the spatial 1st modes are all around 0.44. Most of the other modes of PC change around 0

(Figure 3 (a2–a5 and b2–b5)), and the corresponding absolute value of spatial modes is also less than the first mode. Therefore, from the second PC to the fifth PC, these modes play a less important role in reconstructing data than the first mode, but the later indices show that ignoring these less important modes may lead to less perfect performance of EOF1 reconstruction. In addition, Figure 3 (a1 and b1) illustrates that the amplitudes of PC1 in winter months (November, December, January, and February) were significantly greater than those in summer months (April, May, June, and July). It demonstrates that PM2.5 concentration in winter in North China Plain was significantly higher than that in summer [47].

Figures 4 and 5 depict the four interpolation reconstruction sequences and their residuals for sta 1 and 8, respectively. Both power parameters  $P$  (1 or 2) are adopted for IDW and EOF1 reconstruction for sta 1 and sta 8, but the indices show that choosing  $P=1$  for IDW and EOF1 is more accurate in sta 1, while  $P=2$  for IDW and EOF1 is more accurate in sta 8. The optimal mode number for EOF reconstruction is both three in sta 1 and sta 8. In the part of Result Evaluation and Discussion, we try to explain the reasons for this. It can be seen that four methods can roughly reproduce the valid records in sta 1 and sta 8. In sta 1 (Figure 4), the residuals of the four interpolation methods all change near 0, but there are several errors which are quite different from the observed values. For example, they all show errors of more than  $100 \mu\text{g}/\text{m}^3$  around February 20th and mid-March. Regardless of the instrument failure and other factors, the large error at these times may indicate that the PM2.5 concentration varies greatly among different regions of the same city, and it is not accurate to rely on only the adjacent data in this case. In Figure 5 of sta 8, the situation is similar, but the fluctuation magnitude of the residual sequence is significantly larger than that of sta 1, and the large residuals are also more frequently occurred. The performance of the four methods in sta 8 is generally worse than that of sta 1.

**3.2. Result Evaluation.** In this section, we evaluate four interpolation methods with quantified indices. Figure 6 shows a comparison of 4 interpolation methods in terms of MAE, RMSE, and Corr, and Figure 7 shows bias, NSE, and IA. Because many indices of the TPS method are quite different from those of other methods, in order to see their differences clearly, the indicator values of TPS are directly marked on each subgraph. It can be seen that the EOF1 interpolation performance of sta 1 and sta 8 varies with the number of modes, many indices show that the optimal mode number of EOF1 is three (e.g., Figure 6 (a1 and b1)), and the performance of EOF1 is sometimes worse than other interpolation methods when it is not the optimal mode number. We arrange all six indices of the best performing EOF1 and other three interpolation methods in the descending order of performance. It can be seen that, in sta 1, all 6 indicators show that the performance of EOF1 ( $P=1$ ) is the best (red lines) ( $1\text{-EOF1} > 1\text{-IDW} > \text{SS} > \text{TPS}$ ), while in sta 8, all 6 indicators show that EOF1 ( $P=2$ ) is the best (green

TABLE 2: Distances from station 1 and 8 to other five observation stations and corresponding normalized inverse distance weights when power  $P=1$  and 2, respectively.

| Target station         | Station 1 |        |        |         |        | Station 8 |         |        |        |        |
|------------------------|-----------|--------|--------|---------|--------|-----------|---------|--------|--------|--------|
| Observation stations   | 2         | 3      | 4      | 5       | 6      | 2         | 4       | 5      | 6      | 9      |
| Distance (km)          | 9.19      | 7.62   | 3.86   | 7.52    | 4.43   | 7.88      | 7.48    | 3.96   | 7.36   | 6.07   |
| Weight ( $P=1$ )       | 0.1268    | 0.1530 | 0.3018 | 0.1550  | 0.2634 | 0.1560    | 0.1643  | 0.3105 | 0.1669 | 0.2023 |
| $k$ -th mode ( $P=1$ ) | 0.4471    | 0.1179 | 0.0579 | -0.0574 | 0.0592 | 0.4449    | -0.0714 | 0.0192 | 0.0076 | 0.1143 |
| Weight ( $P=2$ )       | 0.0718    | 0.1045 | 0.4066 | 0.1073  | 0.3098 | 0.1124    | 0.1247  | 0.4453 | 0.1287 | 0.1889 |
| $k$ -th mode ( $P=2$ ) | 0.4474    | 0.2282 | 0.0871 | -0.1222 | 0.1191 | 0.4426    | -0.1291 | 0.0332 | 0.0227 | 0.2540 |
| $G(k)$ (%)             | 98        | 98.78  | 99.35  | 99.72   | 100    | 98.14     | 98.78   | 99.38  | 99.76  | 100    |

It is noteworthy that EOF estimated spatial 1st, 2nd, 3rd, 4th, and 5th modes are listed in the 5th and 7th rows (rather than observation station's coefficients). The contribution of the first  $k$  modes to the total variance  $G(k)$  is listed in the last row.

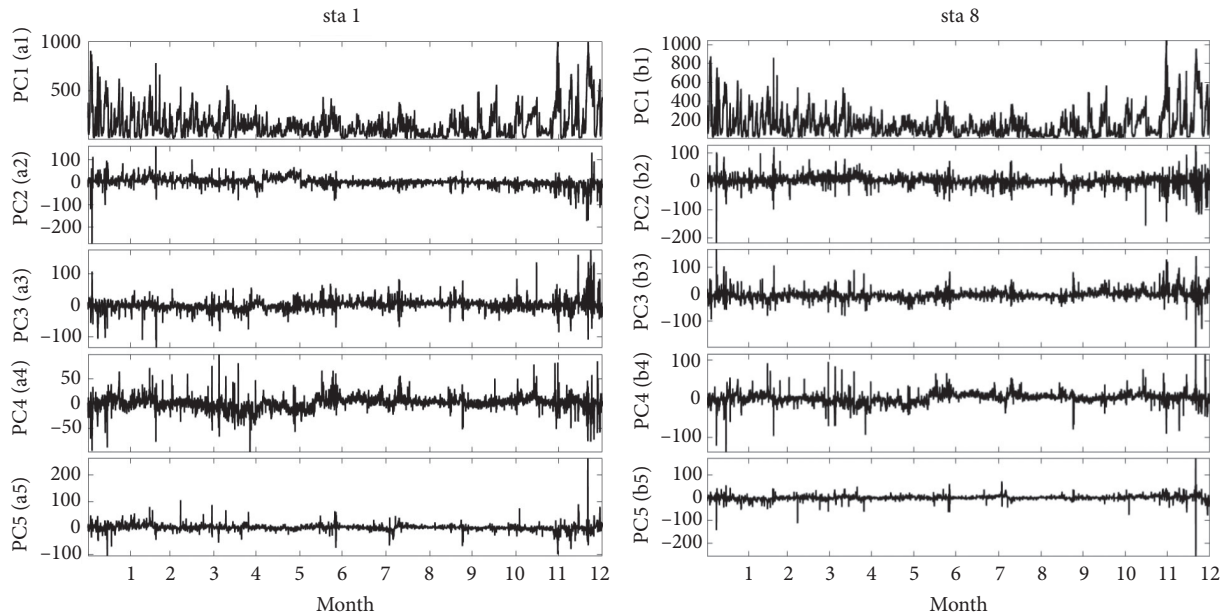


FIGURE 3: Station 1 (a1–a5) and station 8 (b1–b5) temporal variation in EOF modes.

lines) ( $2\text{-EOFI} > 2\text{-IDW} > \text{SS} > \text{TPS}$ ). The IDW performance of many indices is similar; sta 1 prefers  $P=1$ , while sta 8 prefers  $P=2$ . In addition, the indices performance of sta 8 is generally worse than that of sta 1. In Section 4.1, we try to explain why different parameters are chosen at the two sites.

## 4. Discussion

### 4.1. IDW Power $P$ Choice and Sites Number Impact on EOFI.

For the EOFI of this study, we did not take the data of sta 1 and sta 8 into EOF decomposition. The spatial modes of these two stations are calculated by the spatial modes of other 5 stations with IDW, and of course, their spatial modes estimates can also be obtained by other methods, such as Pan and Lv [36] using linear and spline interpolation, respectively, to calculate the spatial modes of river water level measurement points. Next, we try to explain why different  $P$  values are chosen in the two sites as mentioned in Section 3 and discuss the influence of the number of data sites on the EOFI reconstruction.

Firstly, the indices performance of sta 8 is obviously inferior to those of sta 1. There are four same stations (stations 2, 4, 5, and 6) data selected by both sta 1 and sta 8. But the number of missing values at sta 9 for sta 8 imputation is more than that of the sta 3 for sta 1 (the first half of the year missing percent of the sta 9 in Figure 2 reaches 13%), so the completeness of the original data may account for the worse results of sta 8. In addition, for sta 8, when  $P$  is increased from 1 to 2, the EOFI spatial modes and reconstruction sequence will be more dependent on the spatial modes (Table 2) and observation records of the closest station (sta 5), respectively. The adverse impact of the data of sta 9 is reduced, which may be an explanation of sta 8's preference for  $P=2$ .

Furthermore, in previous experiment, data of sta 1 and sta 8 are reconstructed with the data of the other 5 adjacent stations, of which 4 stations (stations 2, 4, 5, and 6) are both used for reconstruction of sta 1 and sta 8. In order to further explore the influence of the remaining station on the interpolation results, another experiment is conducted where the data of sta 3 are not used for sta 1 reconstruction and the

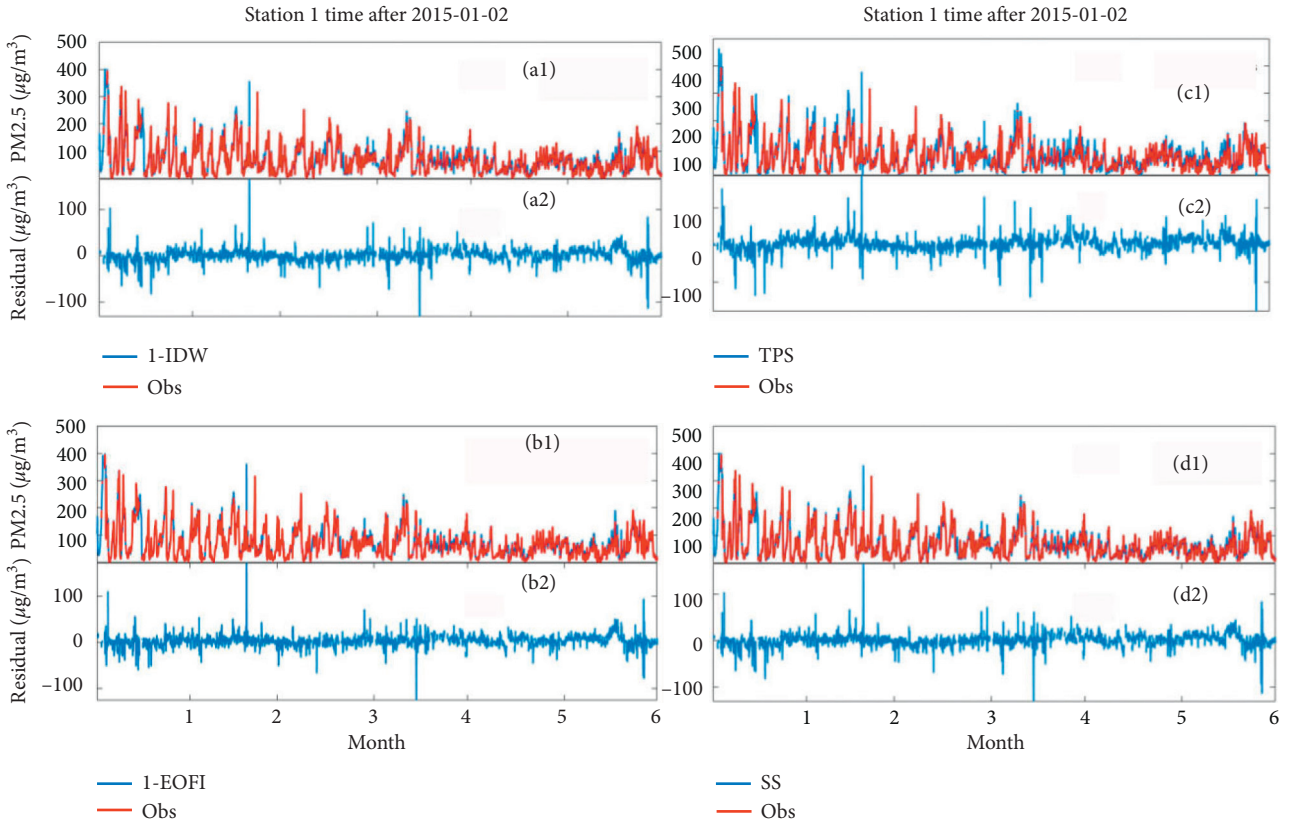


FIGURE 4: 1-IDW (a1),1-EOF1 (b1), TPS (c1), and SS (d1) reconstruction sequence and observations in sta 1 and their corresponding residuals (a2, b2, c2, and d2), 4138 hours after 2015-01-02 1:00. The numbers in front of EOF1 and IDW represent the value of their inverse distance weight  $P$ ; for example, 1-EOF1 represents the value of EOF1's inverse distance weight  $P$  is 1.

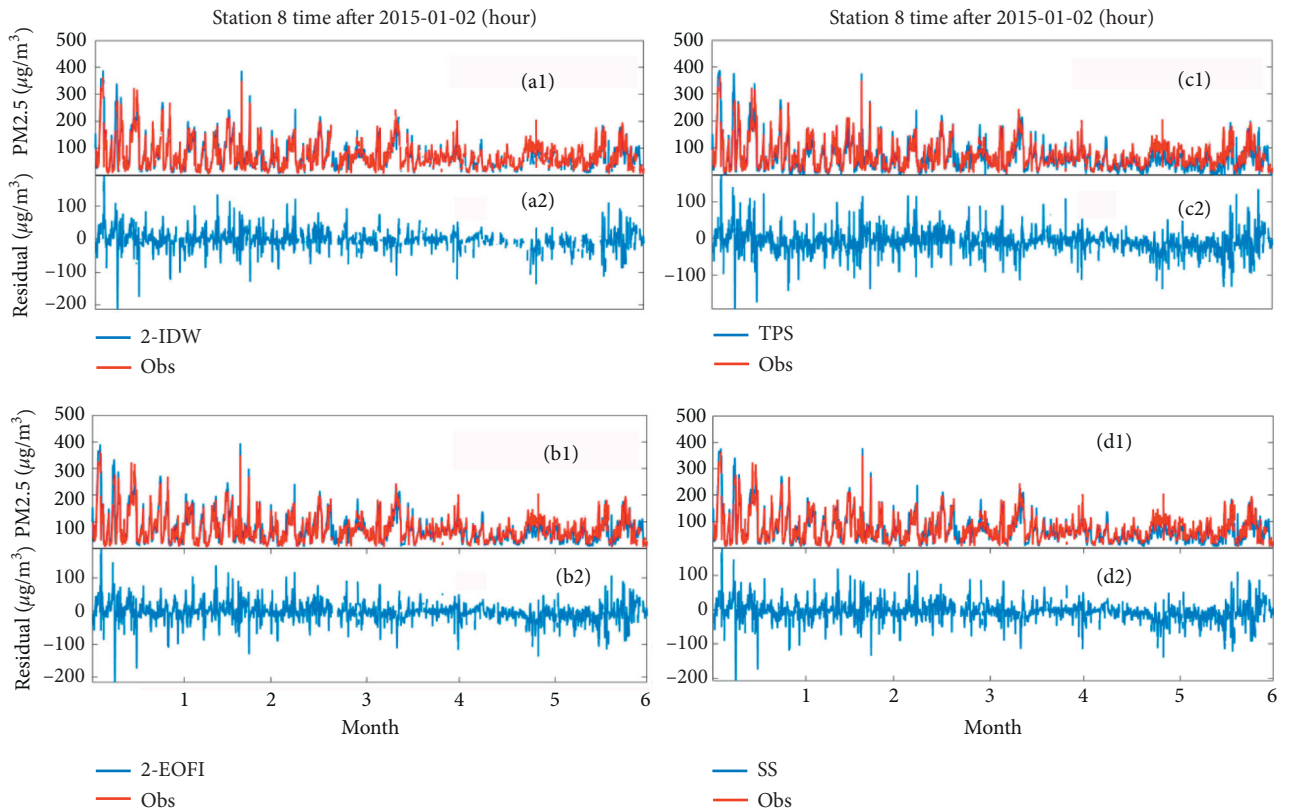


FIGURE 5: Same as Figure 4, but the IDW and EOF1 use  $P=2$  for sta 8.

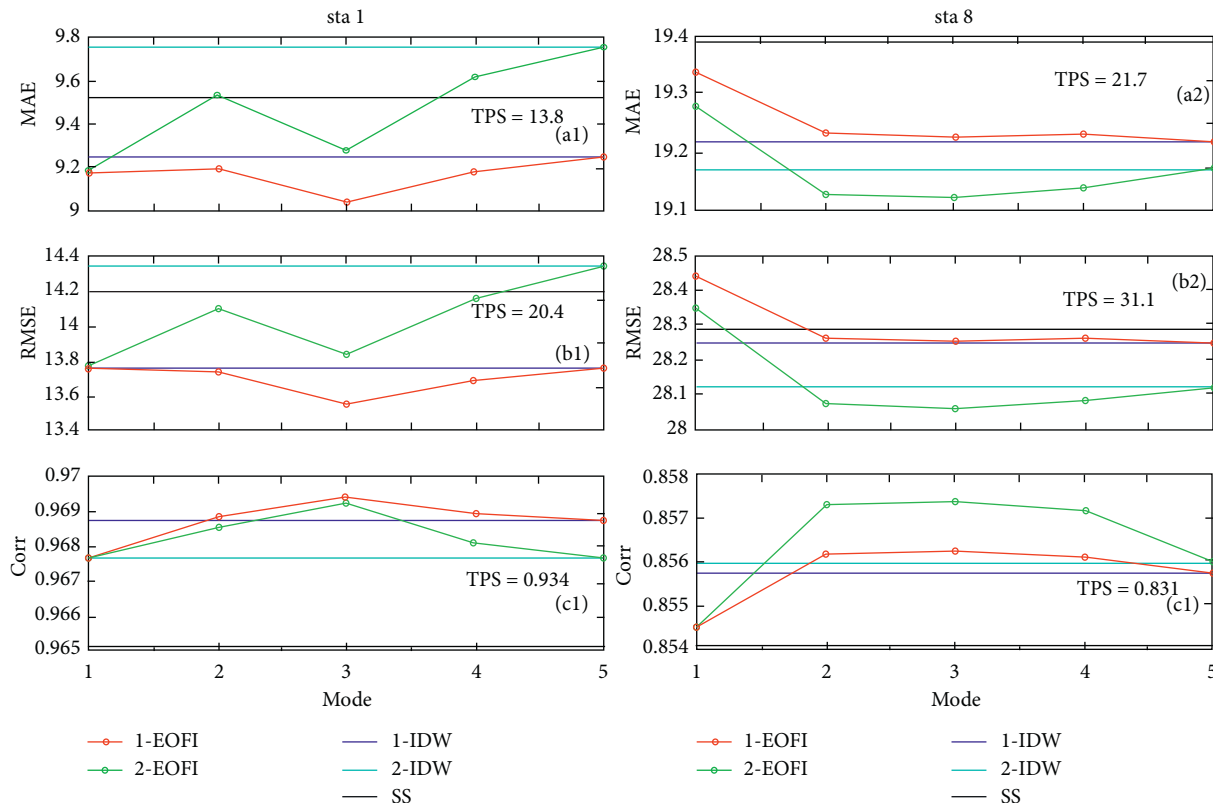


FIGURE 6: MAE, RMSE, and Corr of 4 interpolation methods in sta 1 and sta 8. The horizontal axis is the number of the first  $k$  modes selected in EOFI. The results of the other three interpolation methods are all straight lines and irrelevant to the modes.

data of sta 9 are not used for sta 8. The 4 sites and 5 sites EOFI reconstruction results are shown in Table 3.

It can be seen that, for both sta 1 and sta 8, the EOFI reconstruction with 5 sites is better than that with only 4 sites. In addition, inclusion of data from coastal stations such as sta 10 (Figure 1(b), far away from sta 1 and sta 8) in EOFI is not as good as interpolation with data from only five nearest sites. It is very vital to determine the appropriate number of stations for EOFI according to the feature and quality of the original data. As we can see the performance of using less sites data or adding coastal sites data for EOFI, both of which are worse than that of only five nearest sites data.

**4.2. Further Validation and Impact of Data Time Length on EOFI Results.** In the previous experiment, EOFI selected PM<sub>2.5</sub> data of nearly a full year from five adjacent stations data to perform EOF decomposition and obtained nearly a full year of PC and corresponding spatial modes. In this part, a number of experiments with different lengths of record are implemented to further evaluate and compare the four interpolation methods. Since there are only valid observation records in the first half of 2015 for both sta 1 and sta 8, the reconstruction sequence of four interpolation methods must be compared with valid observations during the same period. Divided by the calendar month, we divided the records in the first half of the year into six one-month sections (Jan, 1; Feb, 2; Mar, 3; Apr, 4; May, 5; and Jun, 6) in

the experimental group E1 and five two-month sections (1-2, 2-3, 3-4, 4-5, and 5-6) in experimental group E2. Four three-month sections (1-3, 2-4, 3-5, and 4-6) are implemented in the experimental group E3. Similarly, E4, E5, and E6 represent the experimental groups with a duration of 4, 5, and 6 months, respectively. There are 21 experiments in total. Since the temporal mode of EOF decomposition is related to the continuity of record, experimental groups with continuous months are set to reduce the inaccuracy of the temporal and spatial modes of EOF decomposition. February in winter and June in summer represents different seasons, and the feature of PM<sub>2.5</sub> concentration is significantly related to the seasons. For example, in winter, more fossil fuels may be consumed for heating; therefore, the PM<sub>2.5</sub> concentration is significantly higher than other seasons.

Figure 8 depicts the main results of EOFI reconstruction sequence of sta 1 and sta 8. It can be seen that, although the spatial 2nd, 3rd, 4th, and 5th modes in different time periods are different, the spatial 1st mode always remains stable at around 0.44, and the corresponding variance contribution also accounts for more than 95% (c1 and c2), which is consistent with the previous results. The RMSE range of EOFI reconstruction for sta 1 is 10–16  $\mu\text{g}/\text{m}^3$  (b1), while the range for sta 8 is 22–36  $\mu\text{g}/\text{m}^3$  (b2). The range is also consistent with the previous results, which shows the stability of the EOFI method. In addition, the number of experiments with the optimal mode number 4 (i.e., using first 4 modes to reconstruct) for sta 1 and sta 8 are both largest,

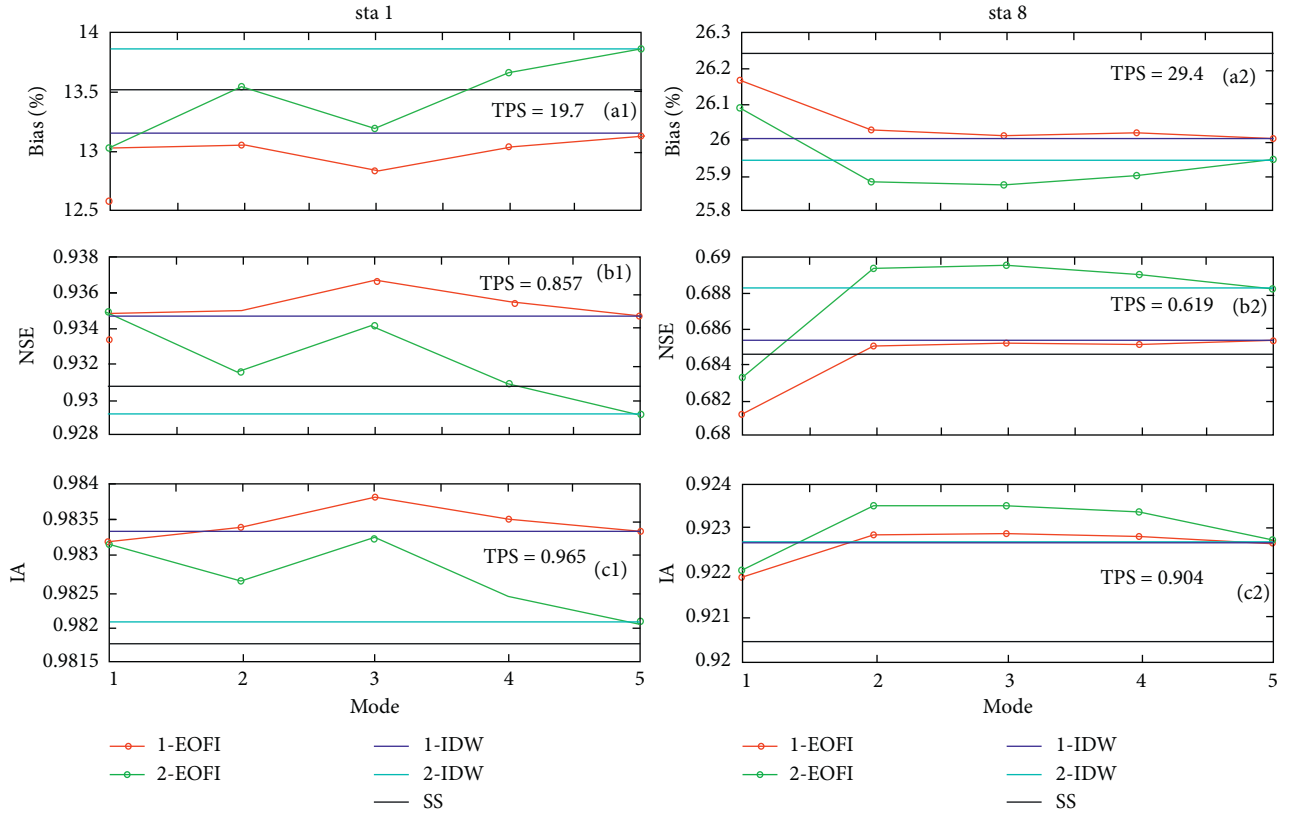


FIGURE 7: Same as Figure 6, but the indices are bias, NSE, and IA in sta 1 and sta 8.

TABLE 3: RMSE between the results of EOF1 reconstruction and the valid observation records of 4 and 5 stations selected by sta 1 and sta 8, respectively.

| RMSE ( $\mu\text{g}/\text{m}^3$ ) | Stations      | Modes  |        |        |        |        |
|-----------------------------------|---------------|--------|--------|--------|--------|--------|
|                                   |               | 1      | 2      | 3      | 4      | 5      |
| sta 1 ( $P=1$ )                   | 2, 3, 4, 5, 6 | 13.759 | 13.739 | 13.558 | 13.686 | 13.764 |
|                                   | 2, 4, 5, 6    | 14.313 | 14.116 | 14.486 | 14.582 | —      |
| sta 8 ( $P=2$ )                   | 2, 4, 5, 6, 9 | 28.344 | 28.072 | 28.061 | 28.082 | 28.119 |
|                                   | 2, 4, 5, 6    | 29.079 | 28.991 | 28.923 | 28.652 | —      |

respectively, but there are still other optimal mode numbers. The optimal mode number can be determined by finding the smallest RMSE [40].

Table 4 compares the performance (in terms of RMSE) of four interpolation methods reconstruction sequence. Among the 21 experiments, there are 19 experiments in sta 1 and 13 experiments in sta 8 showing the RMSE of EOF1 reconstruction is the smallest, respectively. There are also another 7 groups in sta 8 showing SS performed best in terms of RMSE, and these groups mainly include winter months January, February, and March. We infer that this is due to large PM2.5 concentration difference in different sites in winter, and the accuracy of spatial and temporal modes is not as good as those of other seasons.

**4.3. Comparison between EOF1 and DINEOF.** There have been many EOF-based interpolation methods (e.g., DCCEOF in [10], EOF1 in [36], and VE-DINEOF in [40]).

One of the most widely utilized methods is the iterated EOF method, DINEOF [30]. Therefore, it is necessary to compare DINEOF and EOF1 in this study.

First of all, two methods are both based on the matrix eigenvalue decomposition theory, and they all assume that the short missing value intervals of original spatiotemporal observation records will not affect the dominant temporal and spatial modes significantly. Moreover, the first guess values are given to the missing values to enable matrix decomposition. By calculating the RMSE and other indicators, the temporal and spatial modes of the optimal mode number will be used for final reconstruction.

However, the most significant difference between DINEOF and EOF1 is the original data used for matrix decomposition. In EOF1, the data of sta 1 and sta 8 (the second half of the year data is missing) are not included in the decomposed matrix, but in DINEOF, the data of sta 1 and sta 8 are taken into EOF decomposition; firstly, the missing values are replaced with first guess values and then



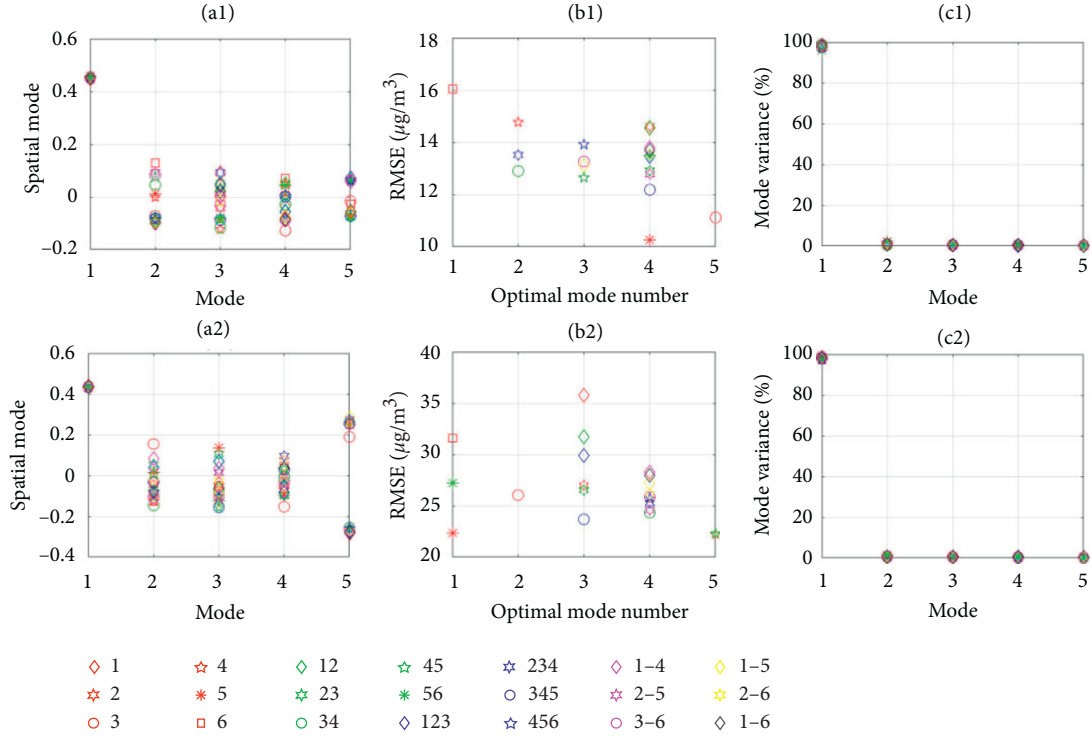


FIGURE 8: EOFI reconstruction sequence's spatial modes, corresponding variance contributions, and the RMSE with optimal mode number of sta 1 (a1, c1, and b1) and sta 8 (a2, c2, and b2) in the first half of the year. The legend color red, green, blue, magenta, yellow, and black represent the experimental groups with durations of 1, 2, 3, 4, 5, and 6 months, respectively.

TABLE 4: RMSE of four interpolation methods in sta 1 and sta 8.

| RMSE Groups | Station Method Month(s) | 1       |         |        |         | 8       |         |        |         |
|-------------|-------------------------|---------|---------|--------|---------|---------|---------|--------|---------|
|             |                         | IDW     | EOF1    | TPS    | SS      | IDW     | EOF1    | TPS    | SS      |
| E1          | 1                       | 14.562  | 14.547* | 23.675 | 15.878  | 36.031  | 35.751  | 38.931 | 35.458* |
|             | 2                       | 14.629  | 14.613  | 20.924 | 13.780* | 27.065  | 26.984  | 28.915 | 26.103* |
|             | 3                       | 11.123* | 11.123* | 13.989 | 11.380  | 26.376  | 26.040  | 27.469 | 25.815* |
|             | 4                       | 15.001  | 14.772* | 22.446 | 15.689  | 22.207* | 22.207* | 24.905 | 22.820  |
|             | 5                       | 10.286  | 10.273* | 16.905 | 10.968  | 22.399  | 22.313* | 27.000 | 24.248  |
|             | 6                       | 16.229  | 16.029* | 23.275 | 16.749  | 32.065  | 31.596* | 36.881 | 33.100  |
| E2          | 1-2                     | 14.596  | 14.557* | 22.308 | 14.840  | 31.889  | 31.688  | 34.317 | 31.159* |
|             | 2-3                     | 12.980  | 12.908  | 17.769 | 12.627* | 26.723  | 26.519  | 28.203 | 25.960* |
|             | 3-4                     | 13.120  | 12.901* | 18.518 | 13.610  | 24.371  | 24.339* | 26.212 | 24.357  |
|             | 4-5                     | 12.732  | 12.650* | 19.716 | 13.406  | 22.303* | 22.303* | 25.965 | 23.539  |
|             | 5-6                     | 13.545  | 13.478* | 20.296 | 14.116  | 27.549  | 27.215* | 32.208 | 28.913  |
| E3          | 1-3                     | 13.507  | 13.459* | 19.855 | 13.753  | 30.175  | 29.934  | 32.211 | 29.497* |
|             | 2-4                     | 13.650  | 13.508* | 19.368 | 13.668  | 25.301  | 25.273  | 27.143 | 24.953* |
|             | 3-5                     | 12.209  | 12.178* | 17.973 | 12.752  | 23.737  | 23.692* | 26.475 | 24.321  |
|             | 4-6                     | 14.016  | 13.913* | 20.991 | 14.625  | 25.846  | 25.758* | 29.911 | 26.984  |
| E4          | 1-4                     | 13.878  | 13.791* | 20.503 | 14.239  | 28.390  | 28.262  | 30.545 | 27.974* |
|             | 2-5                     | 12.858  | 12.791* | 18.757 | 13.018  | 24.614  | 24.594* | 27.108 | 24.780  |
|             | 3-6                     | 13.340  | 13.262* | 19.450 | 13.872  | 25.981  | 25.837* | 29.312 | 26.693  |
| E5          | 1-5                     | 13.204  | 13.114* | 19.800 | 13.616  | 27.310  | 27.196* | 29.878 | 27.278  |
|             | 2-6                     | 13.607  | 13.542* | 19.754 | 13.853  | 26.204  | 26.142* | 29.232 | 26.574  |
| E6          | 1-6                     | 13.764  | 13.691* | 20.432 | 14.197  | 28.119  | 27.985* | 31.096 | 28.283  |

The power  $P$  of IDW and EOF1 is based on the analysis of Section 4.1 (i.e.,  $P = 1$  for sta 1 and  $P = 2$  for sta 8), and the EOF1 reconstruction with optimal mode number is considered. "\*" represents the smallest RMSE of this experiment.

conducted matrix decomposition and iterative replacement until convergence. However, this step may be not suitable for the data processing of a small number of stations because the

first guess values of these missing stations may greatly affect the accuracy of temporal and spatial modes in this case. Even if the final convergent temporal and spatial modes are

obtained through iteration, the calculation resources consumed may be huge. Alvera-Azcárate et al. [32] mentioned that the data points with missing percent more than 95% are removed before data decomposition because they cannot provide effective information. The number of data points involved in their decomposition is huge; therefore, these less-informative points' removal has little impact on the final results. The DINEOF has been widely used for reconstruction of gap-free satellite images where densely sampled and numerous observations are obtained by remote sensing, while in other platforms (e.g., PM2.5 land-based stations in this study and offshore buoy stations array), where observations are relatively rare and sparse sampled, the temporal and spatial modes of iterated EOF methods may be not accurate when there is a large proportion of missing values in the few sites observation data matrix.

Therefore, for the observation records of finite number stations, if we want to make full use of the data of station with large proportion of missing values, EOFI may be more suitable for this kind of interpolation. The superiority of EOFI here is to obtain more reasonable spatial and temporal modes by excluding the records of large missing percent stations before EOF decomposition. All stations share the same time-dependent temporal mode, while the space-dependent spatial mode of the missing data station is estimated by spatial interpolation (IDW is used in this study), and the spatial mode features and patterns are considered. In addition, EOFI can provide more reasonable first guess values for the data of these missing stations, and next, DINEOF is used to iteratively calculate until convergence. For other differences, such as DINEOF iterative decomposition, EOFI can also use iterative decomposition in this study; DINEOF randomly selects a part of observation data as cross validation points, and EOFI here uses the first half year valid observation records and monthly records of sta 1 and sta 8 as check points, both of which can be unified in these aspects.

## 5. Conclusion

In this paper, two-dimensional EOFI is introduced and applied to reconstruct spatial-distributed PM2.5 data as an extension to one-dimensional EOFI in river water level reconstruction. The main step of EOFI here is to calculate the missing data station's estimated spatial modes  $\bar{F}$  by IDW interpolation of spatial modes of the observation sites and then multiply  $\bar{F}$  and the corresponding temporal modes to obtain the EOFI reconstruction sequence, and the optimal mode number of EOFI reconstruction is determined by minimizing RMSE. Compared with the other three interpolation methods (IDW, TPS, and SS), the quantitative indices show that EOFI can improve the interpolation effect. The conclusion is as follows.

TPS and SS have fixed function forms, and their coefficient matrices are space-dependent. The advantage of EOFI is that the spatiotemporal matrix is decomposed into time-dependent temporal modes and space-dependent spatial modes under EOF assumption. Observation stations and missing data stations share the same temporal modes, while the spatial modes of missing data station are estimated by the

IDW of observation stations' spatial modes. The benefit of IDW is that when the distance between the missing station and the observation station is very close, the spatial mode estimated by IDW is very close to that of the observation station; thus, the EOFI reconstruction sequence of the missing station is also close to the data of the observation station, which is consistent with our cognition. More essentially, the IDW weights of neighboring points are generated by statistical estimate of covariance between the observation points. TPS and SS weights do not depend on the statistical features of interpolated fields. EOFI can reduce MAE and RMSE compared with other three methods, and other indices show that the performance of EOFI is better too. This shows that EOFI can improve the interpolation effect with optimal modes. The results of several experimental groups with different data lengths show that the dominant spatial modes of EOF decomposition almost do not change with the time length, which is consistent with the EOF assumption that the spatial modes are independent of time. At the same time, the RMSE of EOFI reconstruction with optimal mode number still shows the advantages over the other three methods.

The proposed method is suitable for interpolation when observations are rare and sparsely distributed, and there are large percent of missing values for some stations' original records. The EOFI reconstruction sequence of missing data station can be a reasonable first guess value for further DINEOF (or other iterated EOF-based method) steps.

EOFI has the advantages of less calculation, less parameter choices, and ease of implementation and can be extended to fill the missing data gaps of other two-dimensional spatial distribution physical variables. The limitation of EOFI is that the missing values' temporal and space gaps should not be too large; otherwise, it will affect the accuracy of spatial and temporal modes. At the same time, the quality of the original data has an impact on the reconstruction results. High quality and complete observation data can produce more accurate spatial and temporal modes, which is conducive to EOFI reconstruction.

## Data Availability

The data (hourly PM2.5 concentration data of 8 stations in Tianjin and station locations) used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The authors would like to thank Professor Yang Gao for providing the PM2.5 concentration data. This work was supported by the National Natural Science Foundation of China (Grant no. 41876003) and the National Key Research and Development Program of China (Grant nos. 2017YFA0604101 and 2016YFC1401404).



## References

- [1] S. Zhai, D. J. Jacob, X. Wang et al., "Fine particulate matter (PM<sub>2.5</sub>) trends in China, 2013–2018: separating contributions from anthropogenic emissions and meteorology," *Atmospheric Chemistry and Physics*, vol. 19, pp. 11031–11041, 2019.
- [2] S. Gautam, A. K. Patra, and P. Kumar, "Status and chemical characteristics of ambient PM<sub>2.5</sub> pollutions in China: a review," *Environment, Development and Sustainability* vol. 21, pp. 1649–1674, 2018.
- [3] H. Shi, S. Wang, J. Li, and L. Zhang, "Modeling the impacts of policy measures on resident's PM<sub>2.5</sub> reduction behavior: an agent-based simulation analysis," *Environmental Geochemistry and Health*, vol. 1, 2019.
- [4] Y. Li, J. Wang, C. Chen, Y. Chen, and J. Li, "Estimating PM<sub>2.5</sub> in the Beijing-tianjin-hebei region using modis aod products from 2014 to 2015," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 41, pp. 721–727, 2016.
- [5] X. Liu and Coauthors, "Fine particulate matter pollution in North China: seasonal-spatial variations, source apportionment, sector and regional transport contributions," *Environmental Research*, vol. 184, Article ID 109368, 2020.
- [6] J. Feng, J. Quan, H. Liao, Y. Li, and X. Zhao, "An air stagnation index to qualify extreme haze events in northern China," *Journal of the Atmospheric Sciences*, vol. 75, pp. 3489–3505, 2018.
- [7] X. Wu, Y. Chen, J. Guo, G. Wang, and Y. Gong, "Spatial concentration, impact factors and prevention-control measures of PM<sub>2.5</sub> pollution in China," *Natural Hazards*, vol. 86, pp. 393–410, 2017.
- [8] L. Zhou, C. Zhou, F. Yang, L. Che, B. Wang, and D. Sun, "Spatio-temporal evolution and the influencing factors of PM<sub>2.5</sub> in China between 2000 and 2015," *J. Geogr. Sci.*, vol. 29, pp. 253–270, 2019.
- [9] P. Yin and Coauthors, "Higher risk of cardiovascular disease associated with smaller size-fractioned particulate matter," *Environmental Science & Technology Letters*, vol. 7, pp. 95–101, 2020.
- [10] K. Bai, K. Li, J. Guo, Y. Yang, and N.-B. Chang, "Filling the gaps of in situ hourly PM<sub>2.5</sub> concentration data with the aid of empirical orthogonal function analysis constrained by diurnal cycles," *Atmospheric Measurement Techniques*, vol. 13, pp. 1213–1226, 2020.
- [11] A. Alvera-Azcárate, A. Barth, D. Sirjacobs, F. Lenartz, and J. M. Beckers, "Data interpolating empirical orthogonal functions (DINEOF): a tool for geophysical data analyses," *Mediterranean Marine Science*, vol. 12, pp. 5–11, 2011.
- [12] D. Kondrashov and M. Ghil, "Spatio-temporal filling of missing points in geophysical data sets," *Nonlinear Processes in Geophysics*, vol. 13, pp. 151–159, 2006.
- [13] J. Elken, M. Zujev, J. She, and P. Lagema, "Reconstruction of large-scale Sea surface temperature and salinity fields using sub-regional EOF patterns from models," *Frontiers Earth Science*, vol. 7, pp. 1–20, 2019.
- [14] L. Feng, G. Nowak, T. J. O. Neill, and A. H. Welsh, "CUTOFF: a spatio-temporal imputation method," *Journal of Hydrology*, vol. 519, pp. 3591–3605, 2014.
- [15] S. Moritz and T. Bartz-Beielstein, "ImputeTS: time series missing value imputation in R," *The R Journal*, vol. 9, pp. 207–218, 2017.
- [16] M. W. Beck, N. Bokde, G. Asencio-Cortés, and K. Kulat, "R package imputetestbench to compare imputation methods for Univariate time series," *The R Journal*, vol. 10, pp. 218–233, 2018.
- [17] N. Bokde, M. W. Beck, F. Martínez Álvarez, and K. Kulat, "A novel imputation methodology for time series based on pattern sequence forecasting," *Pattern Recognition Letters*, vol. 116, pp. 88–96, 2018.
- [18] M. Lepot, J. B. Aubin, and F. H. L. R. Clemens, "Interpolation in time series: an introductory overview of existing methods, their performance criteria and uncertainty assessment," *Water (Switzerland)*, vol. 9, 2017.
- [19] G. Y. Lu and D. W. Wong, "An adaptive inverse-distance weighting spatial interpolation technique," *Computers & Geoscience*, vol. 34, pp. 1044–1055, 2008.
- [20] Y. Chen, X. Shan, X. Jin, T. Yang, F. Dai, and D. Yang, "A comparative study of spatial interpolation methods for determining fishery resources density in the Yellow Sea," *Acta Oceanologica Sinica*, vol. 35, no. 12, pp. 65–72, 2016.
- [21] X. Zong, M. Xu, J. Xu, and X. Lv, "Improvement of the ocean pollutant transport model by using the surface spline interpolation," *Tellus A: Dynamic Meteorology and Oceanography*, vol. 70, pp. 1–13, 2018.
- [22] G. P. Cressman, "An operational objective analysis system," *Monthly Weather Review*, vol. 87, pp. 367–374, 1959.
- [23] S. Chen, C. F. N. Cowan, and P. M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 2, pp. 302–309, 1991.
- [24] J. P. C. Kleijnen, "Kriging metamodeling in simulation: a review," *European Journal of Operational Research*, vol. 192, pp. 707–716, 2009.
- [25] Y. C. Fang, T. J. Weingartner, R. A. Potter, P. R. Winsor, and H. Statscewich, "Quality assessment of HF radar-derived surface currents using optimal interpolation," *Journal of Atmospheric and Oceanic Technology*, vol. 32, pp. 282–296, 2015.
- [26] Z. H. Liu, R. G. Huang, Y. M. Hu, S. D. Fan, and P. H. Feng, "Generating high spatiotemporal resolution LAI based on MODIS/GF-1 data and combined Kriging-Cressman interpolation," *International Journal of Agricultural and Biological Engineering*, vol. 9, pp. 120–131, 2016.
- [27] G. Burgers, P. J. Van Leeuwen, and G. Evensen, "Analysis scheme in the ensemble Kalman filter," *Monthly Weather Review*, vol. 126, pp. 1719–1724, doi. 10.1175/1520-1998)126 2.0.CO;2 1998.
- [28] T. M. Smith, R. W. Reynolds, R. E. Livezey, and D. C. Stokes, "Reconstruction of historical Sea surface temperatures using empirical orthogonal functions," *Journal of Climate*, vol. 9, pp. 1403–1420, 1996.
- [29] K. Y. Kim, "Statistical interpolation using cyclostationary EOFs," *Journal of Climate*, vol. 10, pp. 2931–2942, 1997.
- [30] J.-M. Beckers and M. Rixen, "EOF calculations and data filling from Incomplete Oceanographic Datasets," *Journal of Atmospheric and Oceanic Technology*, vol. 20, pp. 1839–1856, 2003.
- [31] C. Jayaram, N. Priyadarshi, J. Pavan Kumar, T. V. S. Udaya Bhaskar, D. Raju, and A. J. Kochuparampil, "Analysis of gap-free chlorophyll-a data from MODIS in Arabian Sea, reconstructed using DINEOF," *International Journal of Remote Sensing*, vol. 39, pp. 7506–7522, 2018.
- [32] A. Alvera-Azcárate, A. Barth, M. Rixen, and J. M. Beckers, "Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: Application to the Adriatic Sea surface temperature," *Ocean Model*, vol. 9, pp. 325–346, 2005.
- [33] Y. C. Liang, M. R. Mazloff, I. Rosso, S. W. Fang, and J. Y. Yu, "A multivariate empirical orthogonal function method to

- construct nitrate maps in the Southern Ocean,” *Journal of Atmospheric and Oceanic Technology*, vol. 35, pp. 1505–1519, 2018.
- [34] Z. Zhang, X. Yang, H. Li, W. Li, H. Yan, and F. Shi, “Application of a novel hybrid method for spatiotemporal data imputation: a case study of the Minqin County groundwater level,” *Journal of Hydrology*, vol. 553, pp. 384–397, 2017.
- [35] D. Sirjacobs, A. Alvera-Azcárate, A. Barth et al., “Cloud filling of ocean colour and sea surface temperature remote sensing products over the Southern North Sea by the Data Interpolating Empirical Orthogonal Functions methodology,” *Journal of Sea Research*, vol. 65, pp. 114–130, 2011.
- [36] H. Pan and X. Lv, “Reconstruction of spatially continuous water levels in the Columbia River estuary: the method of empirical orthogonal function revisited,” *Estuarine, Coastal and Shelf Science*, vol. 222, pp. 81–90, 2019.
- [37] P. Matte, D. A. Jay, and E. D. Zaron, “Adaptation of classical tidal harmonic analysis to nonstationary tides, with application to river tides,” *Journal of Atmospheric and Oceanic Technology*, vol. 30, no. 3, pp. 569–589, 2013.
- [38] J. Li and A. D. Heap, “A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors,” *Ecological Informatics*, vol. 6, pp. 228–241, 2011.
- [39] E. N. Lorenz, *Empirical Orthogonal Functions and Statistical Weather Prediction*, Massachusetts Institute of Technology, Cambridge, MA, USA, 1956.
- [40] B. Ping, F. Su, and Y. Meng, “An improved DINEOF algorithm for filling missing values in spatio-temporal sea surface temperature data,” *PLoS One*, vol. 11, pp. 1–12, Article ID e0155928, 2016.
- [41] W. R. Tobler, “A computer movie simulating urban growth in the Detroit region,” *Journal of Economic Geography*, vol. 46, pp. 234–240, 1970.
- [42] J. Duchon, “Splines minimizing rotation-invariant seminorms in Sobolev spaces,” in *Constructive Theory of Functions of Several Variables*, pp. 85–100, Springer, Berlin, Germany, 1977.
- [43] F. L. Bookstein, “Principal Warps: thin-plate splines and the decomposition of deformations,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 567–585, 1989.
- [44] Z. Guo, H. Pan, W. Fan, and X. Lv, “Application of surface spline interpolation in inversion of bottom friction coefficients,” *Journal of Atmospheric and Oceanic Technology*, vol. 34, pp. 2021–2028, 2017.
- [45] J. E. Nash and J. V. Sutcliffe, “River flow forecasting through conceptual models. Part 1 — a discussion of principles,” *Journal of Hydrology*, vol. 10, pp. 282–290, 1970.
- [46] C. J. Willmott, “On the validation of models,” *Progress in Physical Geography*, vol. 2, pp. 184–194, 1981.
- [47] X. Wang, R. R. E. Dickinson, L. Su, C. Zhou, and K. Wang, “PM 2.5 pollution in China and how it has been exacerbated by terrain and meteorological conditions,” *Bulletin of the American Meteorological Society*, vol. 99, pp. 105–120, 2018.

## Research Article

# Research on the Measurement, Evolution, and Driving Factors of Green Innovation Efficiency in Yangtze River Economic Belt: A Super-SBM and Spatial Durbin Model

Renyan Long <sup>1</sup>, Hangyuan Guo,<sup>2</sup> Danting Zheng,<sup>2,3</sup> Ronghua Chang,<sup>2,4</sup> and Sanggyun Na <sup>2</sup>

<sup>1</sup>School of Economics and Management, Xinyu University, Xinyu 338004, China

<sup>2</sup>School of Business Administration, Wonkwang University, 460 Iksandae-ro, Iksan, 54538, Republic of Korea

<sup>3</sup>School of Economics and Management, Shangrao Normal University, Shangrao 334001, China

<sup>4</sup>School of Economics and Management, Shanxi Datong University, Datong 037009, China

Correspondence should be addressed to Sanggyun Na; [nsghy@wku.ac.kr](mailto:nsghy@wku.ac.kr)

Received 28 July 2020; Revised 6 September 2020; Accepted 19 September 2020; Published 27 October 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Renyan Long et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the shortage of resources and the increasingly serious environmental pollution in China, green innovation has become a sustainable competition for a region. The Yangtze River Economic Belt (YREB) strategy is one of the most important strategies for the sustainable development of China's economy under the new normal. Green innovation plays a linking role in the resources exchange and trade flow in YREB, and it is also the foundation and guarantee to implement the YREB strategy. The global environmental pollution and the weak recovery of world economy make the traditional extensive economic growth model unsustainable. Sustainable economic growth should focus on the quality of development and its external costs to the environment. In order to implement the concept of sustainable development, the improvement of logistics ecological efficiency is related to the quality of ecological civilization construction. Therefore, it is of theoretical and practical significance to study the measurement, evolution, and driving factors of coordinated development level of regional green innovation system. This paper proposes a super-slack-based measure (super-SBM) data envelopment analysis (DEA) model to measure the green innovation efficiency of 11 provinces and cities in YREB from 2008 to 2017, mastering its spatial and evolutionary characteristics, and conduct empirical analysis on the influencing factors. The empirical results indicate that economic development, government support, and industrial structure upgrading are the leading forces to directly enhance the green technology innovation ability of cities in the Yangtze River Economic Belt and play the core driving role of green innovation. To further enhance the capacity of urban green innovation in the Yangtze River Economic Belt, we will increase the government's support for green innovation, optimize the environmental governance model, promote the green upgrading of industrial structure, and enhance the enthusiasm of enterprises for green innovation.

## 1. Introduction

With the adoption of China's reform and open up policy, the miracle of rapid economic growth is attracting the world's attention. In the meantime, the emerging issues of environment and resource depletion also pose a huge challenge to economic development. Coordinating the relationship between economy and environment is the key to the implementation of sustainable development strategy [1], made in China 2025, and raised the "green innovation-

driven development" strategy to promote the healthy development of national economy.

As an important support belt for China's economic development in the new era, the YREB spans three regions in China, connects the Yangtze River Delta Basin with the most developed economy in China, and together with the coastal economic belt forms the T-shaped model of economy development in China, and it is important in promoting and demonstrating industrial transformation and upgrading and green development. With the diversification of innovation

sources, ecological environment has become an important factor in innovation transformation and upgrading. However, there are many industries with high pollution and high-energy consumption in the YREB. Environmental pollution and resource shortage have become important factors restricting the YREB strategic development. The contradiction of ecological environment and economic society has become extremely acute. Under the rigid constraints of energy and environment, it is urgent to incorporate green development concept into technological innovation [2]. Therefore, as the integration point of innovation-driven and green development, green innovation has become an effective means to break through the constraints of resources and environment and promote sustainable development. It is of great practical significance to improve the green innovation efficiency of the YREB and realize the win-win of innovation efficiency, ecological efficiency, and economic efficiency, so as to enhance the ability of regional sustainable development and promote high-quality economic development [3].

In early 2016, Present Xi Jinping proposed that the ecological environment of the YREB should be placed in an overwhelming position. In March 2016, the meeting of the Political Bureau of the CPC Central Committee presided over by General Secretary Xi approved the outline of the development plan of the Yangtze River Economic Belt, stressing once again that “development should be promoted on the premise of protecting the ecology, enhancing the overall planning, integrity, coordination, and sustainability of development, and improving the efficiency of essential allocation.” In March 2018, it was clearly pointed out in the work report of the 19<sup>th</sup> National Congress of the Communist Party of China that the development orientation of the Yangtze River Economic Belt should be guided by ecological priority and green development, promote the overall layout of “five in one:” innovation, coordination, green, opening, and sharing, accelerate the construction of ecological civilization, and achieve regional coordination and sustainable development. The fundamental way to adjust the regional economic structure, change the mode of economic development, and promote high-quality economic development lies in the continuous green innovation. “Green innovation” is a complex system including resource input, innovation output, and environmental benefits. It coordinates the relationship between economic development and ecological protection, realizes the optimal benefit output with the least resource input, and creates the highest green innovation efficiency.

This paper modifies the super-SBM-DEA model to measure the green innovation efficiency of 11 provinces and cities in YREB from 2008 to 2017, mastering its spatial and evolutionary characteristics, and conduct empirical analysis on the influencing factors. The main contribution is using the undesirable super-SBM-DEA method to avoid any underestimation or overestimation of the green innovation efficiency caused by radial and nonradial DEA. The remainder of this paper is organized as follows. The literature review is presented in Section 2. Section 3 briefly describes the measurement method for green innovation efficiency.

Section 4 presents the data and variables. Empirical results and analysis are reported in Section 5. Section 6 draws conclusions and policy implications. The article structure is shown in Figure 1.

## 2. Literature Review

According to the needs of balanced development in ecology and economy, we need to find a relationship among the rapid development of economy, excessive use of resources, and deterioration of natural environment. By measuring green innovation efficiency, we can find key influencing factors and promote the sustainable development of green ecological economy [4–6]. According to existing literatures, research on green innovation efficiency can be summarized in the following three aspects: (1) research on green innovation efficiency; (2) research on the measurement of green innovation efficiency; (3) research on the influencing factors of green innovation efficiency.

*2.1. Research on Green Innovation Efficiency.* Green innovation has become a popular concept, and it is often known as ecological innovation, sustainable innovation, and environmental innovation [7]. Fussier and James first introduced the term green innovation in the book driving green innovation, defining as new products or processes which provide customer and business value but significantly decrease environmental impacts [8]. Kemp et al. define green innovation as a new process technology, system, and product to avoid or reduce environmental damage [9]. Compared with traditional innovation, green innovation takes both economic and environmental benefits into account and adapts to the improvement of supply side structural reform quality and efficiency of industrial parks. From the perspective of systems theory, green innovation is a combination of industrial innovation system theory and green economy theory, in reference to both green products and green processes [10, 11], including the introduction of any new or significantly improved product, process, organizational change, or marketing solution to reduce the consumption of natural resources and the emission of harmful substances in the product life cycle [12].

In 1951, Kaufman first put forward the concept of “efficiency.” He pointed out that if technology cannot realize the increase or decrease in output or input at the given level of output or input, the input-output vector in this state was defined as technology efficiency. Then, Schumpeter combines the concept of innovation and efficiency and points out that the fundamental purpose of innovation is to maximize regional economic and social benefits. Feng Zhijun defined green innovation efficiency as an input-output efficiency that can promote the unity of “economic benefits, environmental benefits, and social benefits” [13]. The authors in [14] pointed out that green innovation efficiency should not only reflect “green” and “innovation” but also reflect its economic characteristics, that is, economic efficiency. In addition, the authors in [15] believe that green innovation efficiency is a

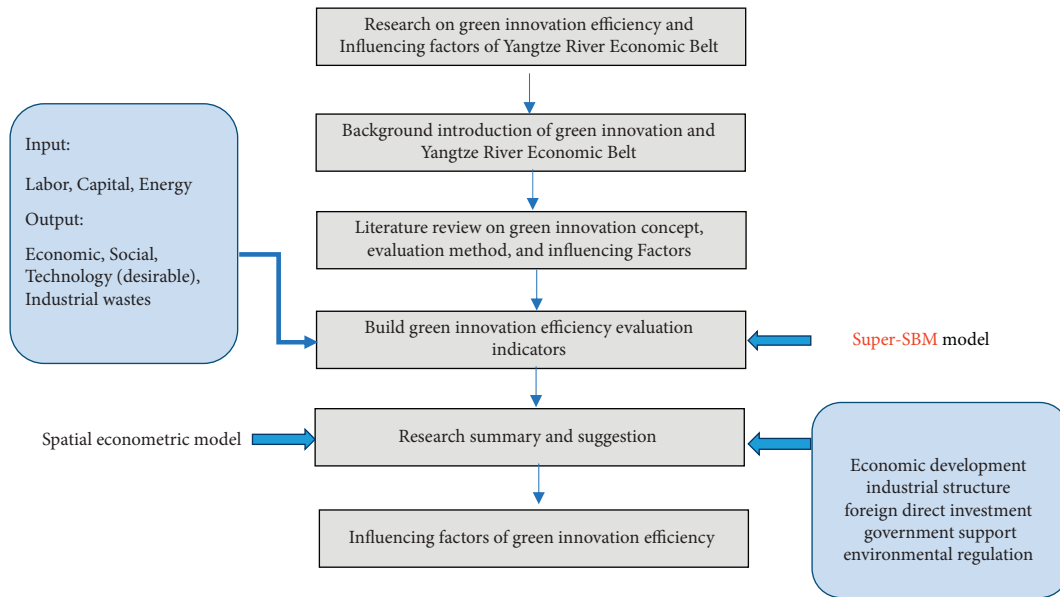


FIGURE 1: Article structure.

comprehensive innovation efficiency considering the cost of resource consumption and environmental pollution.

**2.2. Research on the Measurement of Green Innovation Efficiency.** There are two main methods to measure green innovation efficiency: the parametric analysis method, using SFA, and nonparametric analysis method, using DEA [16, 17]. Parametric analysis assumes that the departure from the frontier of the DMU is the result of a combination of stochastic disturbances and technical inefficiencies. The application of SFA focused mainly on enterprises' efficiency and its influencing factors and on research in the economic field [18, 19]. The authors in [20] measured the green innovation efficiency in China's provinces based on the improved stochastic frontier model and demonstrated the spatial agglomeration characteristics and path dependence of interprovincial green innovation efficiency from a spatial perspective.

The nonparametric analysis method constructs a minimum output possibility set that can accommodate all individual production modes according to the input and output of all decision-making units in the sample and measures the input-output efficiency based on the production possibility set. The authors in [21] compared the innovation efficiency of 185 regions in 23 European countries with the multiobjective DEA model and pointed out that there were differences in the innovation efficiency between different regions and different innovation stages. The authors in [22] used DEA to measure the innovation efficiency values of hospitals in 29 OECD countries between 2000 and 2010 and then applied the panel Tobit model to determine the environmental factors affecting hospital efficiency scores. By decomposing the Malmquist Productivity Index Decomposition, the change in the efficiency decomposition value was analyzed. The authors in [23] used the DEA method to calculate the overall efficiency, patent

production efficiency, and scientific paper production efficiency of 32 Mexican states. The authors in [24] used the SBM model to measure the green innovation efficiency of Chinese industrial enterprises without considering the nonexpected output and analyzed the regional differences of the green innovation efficiency of industrial enterprises in the regions. Luo et al. [3] applied the Malmquist Index and data envelopment analysis to evaluate the efficiency of green technology innovation in strategic emerging industries. Du et al. [25] used a two-stage network DEA with shared input to measure the efficiency of regional enterprises' green technology innovation and explored the regional differences in industrial enterprises' green technology R&D and the efficiency of green technology achievement transformation.

**2.3. Research on the Influencing Factors of Green Innovation Efficiency.** The influencing factors of green innovation efficiency can be classified into direct factors and indirect factors. Direct factors included labor quality, industrial structure, resource consumption, and technological innovation. The authors in [26, 27] conducted a dynamic evaluation on the efficiency of technological innovations in OECD countries and 20 member states of the European Union based on the Malmquist Index. Guan and Zuo [28] applied dual network DEA model to compare technological innovation efficiency of 35 countries. Yu et al. [29] considered the direct factors such as human capital, enterprise nature, and industrial structure when measuring the efficiency of technological innovation in China and found that they all have a significant impact on the efficiency of technological innovation. Wang et al. [30] found that R&D investment intensity has double threshold effect on green innovation efficiency of high-tech industry based on provincial panel data from 2006 to 2012.

The indirect factors included economic development, government funding, regional infrastructure, foreign direct investment, opening up, and environmental regulation. Yu et al. [31] reveal that environmental regulation can significantly improve the green innovation efficiency of the YREB, but different environmental regulation models have different effects on the green innovation efficiency of the YREB. Yang et al. [32] further analyzed the driving mechanism of green innovation efficiency in the YREB and found that the cost of enterprise pollution, the maturity of technology market, and the openness of market are conducive to promoting the efficiency of green innovation in the YREB, while the industrial structure has no significant impact on the efficiency of green innovation. Luo et al. [3] revealed the impact of international R&D capital technology spillover on the efficiency of green technology innovation by building a spatial model. According to the agglomeration effect of FDI, Gong et al. [33] demonstrated the effect and transmission mechanism of industrial green innovation efficiency.

According to the existing literatures on green innovation, there are still some shortcomings: Renyan Long concentrated on the level of enterprises, industrial, or provinces. The research on the efficiency of regional green innovation or specific economic zones and economic basins is of great practical significance for promoting the coordinated development of regional economy. At the same time, factors flow, technology spillover, pollution discharge, and innovation all show spatial interdependence and mutual influence. Existing literatures ignore the spatial correlation of green innovation as data with spatial correlation. Using the 11 provinces/cities' panel data in the YREB from 2008 to 2017, this paper uses the super-slack-based measure (super-SBM) model to evaluate the green innovation efficiency, using Global Moran Index to analyze the spatial correlation and spatial agglomeration characteristics and finally use the spatial measurement model to analyze the influencing factors of green innovation efficiency.

### 3. Materials and Methods

**3.1. Super-SBM Model.** Traditional DEA models, such as the CCR and BCC models, are radial projection constructs by Cook and Seiford [34], which assumes that all the outputs of a production system are valuable and should be maximized for given inputs. Nevertheless, the undesirable output will have significant effects on the efficiency in the whole process [35, 36]. Tone [37] developed a nonradial measurement to solve the problems of input and output slacks by proposing the slack-based measure (SBM). Compared with the traditional DEA, the efficiency value of this method is distributed in the (0, 1) interval, and the efficiency value of the effective DMU is 1. Therefore, when there are multiple effective DMUs, further comparison cannot be made. Then, Tone [38] developed a superefficiency SBM-DEA model which solves the problem of effective sorting and allows the efficiency score to be greater than 1 and can be easily rank-efficient DMUs. Super-SBM model can not only deal with the unexpected output more appropriately but also make further comparison in effective decision-making units, so it is more accurate and

rigorous. Therefore, the super-SBM with undesirable outputs is introduced into measuring green innovation efficiency in this study, and the model is as follows:

$$\min \rho^* = \frac{(1/m) \sum_{i=1}^m (\bar{x}_i/x_{i0})}{(1/(s_1 + s_2)) \left( \sum_{r=1}^{s_1} (\bar{y}_r^g/y_{r0}^g) + \sum_{r=1}^{s_2} (\bar{y}_r^b/y_{r0}^b) \right)}, \quad (1)$$

$$\text{s.t.} \begin{cases} \bar{x} \geq \sum_{j=1, \neq 0}^n \lambda_j x_j, \\ \bar{y}^g \leq \sum_{j=1, \neq 0}^n \lambda_j y_j^g, \\ \bar{y}^b \geq \sum_{j=1, \neq 0}^n \lambda_j y_j^b, \\ \bar{x} \geq x_0, \quad \bar{y}^g \leq y_0^g, \quad \bar{y}^b \geq y_0^b, \quad \lambda \geq 0, \end{cases} \quad (2)$$

where  $\rho^*$  is the green innovation efficiency and  $\lambda$  is the constant vector. The super-SBM model is simultaneously able to measure DMU efficiency and can also calculate DMU input and undesirable output redundancy rates, and it fully considers and effectively solves the problem with undesirable output and is more accurate to evaluate and analyze regional sustainable development.

### 3.2. Spatial Econometric Model

**3.2.1. Spatial Autocorrelation Analysis.** Spatial autocorrelation analysis is a kind of spatial data analysis method that is used for the estimation and analysis of spatial dependency and heterogeneity among objects, which is commonly indicated by Moran Index (Moran's I) [39–42]. Before using spatial econometric methods, it is needed to be constructed to examine whether the green innovation efficiency in YREB has spatial dependence.

Global spatial autocorrelation is used to measure the distribution characteristics of the entire research unit among spatial elements, and it can effectively test the autocorrelation of adjacent units. The global Moran's I value ranges from  $[-1, 1]$ . If  $I < 0$ , there is a negative spatial correlation, which indicates that the efficiency in the study area is in a discrete state. If  $I > 0$ , there is a positive correlation, indicating an agglomeration state. If  $I = 0$ , demonstration is made that the treatment efficiency is random, and the formula is as follows:

$$\text{Moran's } I = \frac{\sum_{i=1}^n \sum_{j \neq 1}^n W_{ij} (x_i - \bar{x})(x_j - \bar{x})}{S^2 \sum_{i=1}^n \sum_{j \neq 1}^n W_{ij}}, \quad (3)$$

where  $S^2 = (1/n) \sum_{i=1}^n (x_i - \bar{x})^2$ ,  $\bar{x} = (1/n) \sum_{i=1}^n x_i$ ,  $S^2$  is the variance value of green innovation efficiency,  $n$  represents the total 11 provinces/cities in YREB,  $x_i$  and  $x_j$  show province  $i$  and province  $j$ 's green innovation efficiency,  $\bar{x}$  represents the average green innovation efficiency, and  $w_{ij}$  is the spatial weighting matrix.



**3.2.2. Spatial Weighting Matrix.** Setting spatial weighting matrix is the basis of the spatial autocorrelation test and spatial econometric model. It reflects the spatial distance between two regions, usually including geographical distance and socioeconomic distance. At present, geographic distance is more common in research. The spherical distance ( $d$ ) between provincial capitals can be used to construct the spatial weighting matrix of geographical distance [43]. It uses the reciprocal of the square of the central distance between regions. The specific formula is as follows:

$$W_{ij} = \begin{cases} \frac{1}{d_{ij}^2}, & (i \neq j), \\ 0, & (i = j). \end{cases} \quad (4)$$

**3.2.3. Spatial Econometric Model.** Following Elhorst and Geogr [44], there are mainly three kinds of spatial econometric models: spatial lag panel model (SLM), spatial error panel model (SEM), and spatial Durbin panel model (SDM). The SLM model hypothesizes that the value of the dependent variable observed at a particular location is partially determined by a spatially weighted average of neighboring-dependent variables.

If the level of green innovation efficiency in the region is not only affected by some variables in the region and by the level of green innovation efficiency in neighboring regions, the spatial lag model (SLM) can be set up, which can be expressed as follows:

$$\ln \text{GIE}_{it} = \alpha_{it} + \rho W \ln \text{GIE}_{it} + X_{it}\beta_{it} + \mu_{it}, \quad (5)$$

where  $\alpha$  is the constant term and  $W$  is the spatial weighting matrix.  $X$  is the variable matrix of the corresponding influencing factors after the logarithmic treatment, and  $\beta$  is the influencing coefficient of the local influencing factors on the local green innovation efficiency.  $i$  represents the corresponding region,  $t$  represents the corresponding year, and  $\mu$  is the random error term.  $\rho$  is the spatial lag variable influence coefficient of green innovation efficiency development, which reflects the spillover effect of green innovation efficiency development on green innovation development in the surrounding areas of the target area.

If the spatial dependence of green innovation behavior is affected by some error disturbance terms which are difficult to observe and have certain spatial structure, and to effectively measure the impact of this error impact on the efficiency of green innovation in this region, the spatial error model (SEM) can be expressed as follows:

$$\begin{aligned} \ln \text{GIE}_{it} &= \alpha_{it} + \rho W \ln \text{GIE}_{it} + X_{it}\beta_{it} + \mu_{it}, \\ \mu_{it} &= \lambda W \mu_{it} + \varepsilon_{it}, \end{aligned} \quad (6)$$

where the parameter  $\lambda$  reflects the regional spillover effects caused by the error term and  $\varepsilon$  is the residual term.

If the level of green innovation efficiency in the region is not only affected by the spatial spillover effect of green innovation efficiency in neighboring regions but also by other variables in neighboring regions, the spatial Durbin model (SDM) can be considered, which can be expressed as follows:

$$\ln \text{GIE}_{it} = \alpha_{it} + X_{it}\beta_{it} + W X_{it}\theta_{it} + \mu_{it}, \quad (7)$$

where  $\theta$  reflects the weighted influence of other regional factors on the efficiency of green innovation in this region, which is defined as other spillover effect in this paper.

**3.2.4. Decomposition of Direct and Indirect Effects.** Due to the spatial correlation in the spatial regression models, the authors in [45] point out that the coefficients of the explanatory variables in the regression model cannot accurately reflect the marginal effect. Spatial spillover effect is an important analysis tool in the spatial econometric model. Because spillover effect has a certain direction of source and source, there will be other spillover effects of other regional influencing factors on innovation efficiency in the region, and there will be other spillover effects of regional relevant variables on green innovation efficiency in the surrounding regions. In the spatial econometric model, the independent variable and the dependent variable will interact. At this time, the marginal effect of the independent variable on the dependent variable cannot be regressed by the linear model. Further deconstruction is needed to simplify the above spatial Durbin model into a vector expression at a specific time point:

$$\ln \text{GIE}_{it} = (1 - \rho W)^{-1} \alpha y_N + (1 - \rho W)^{-1} \cdot (\beta \ln X_i + \theta W \ln X_i) \mu^*, \quad (8)$$

where  $y_N$  is the vector of  $N \times 1$ -order dependent variable,  $\alpha$  is the constant term,  $\mu^*$  is the cross-section, random, and period error term, and  $\ln X_i$  is the  $N \times K$  dimension matrix composed of all independent variables. At a specific time point, the derivative matrix expression of the dependent variable  $\ln \text{GIE}_{it}$  to the independent variable  $K$  is

$$\begin{aligned} \left[ \frac{\partial \ln \text{GIE}}{\partial \ln x_1^k}, \dots, \frac{\partial \ln \text{GIE}}{\partial \ln x_N^k} \right] &= (1 - \rho W)^{-1}, \\ \begin{bmatrix} \beta_k & W_{12}\theta_k & \dots & W_{1N}\theta_k \\ W_{21}\theta_k & \beta_k & \dots & W_{2N}\theta_k \\ \vdots & \vdots & & \vdots \\ W_{N1}\theta_k & W_{N2}\theta_k & & \beta_k \end{bmatrix} &. \end{aligned} \quad (9)$$

The mean value of the elements on the main diagonal of the right matrix in the formula reflects the influence degree of the independent variable on the dependent variable in the province, that is, the effect of a province on the efficiency of

green technology innovation in the region through a certain influencing factor, which is called direct spillover effect. It is expressed as  $\overline{M}_{\text{direct}} = N^{-1} \cdot \text{tra}[X(W)]$ , where  $\text{tra}[X(W)]$  is the trace of matrix  $X(W)$ , and it is the sum of main diagonals. The mean value of other elements on the nonmain diagonal of the right matrix in this formula reflects the spillover effect of a province on the green innovation efficiency of other provinces through its own relevant influencing factors, which is called the indirect spillover effect [46]. In this paper, it is defined as the spillover effect; that is,  $N^{-1} \cdot y \cdot X(W) \cdot y - N^{-1} \cdot \text{tra}[X(W)]$ . Finally, direct spillover effect and indirect spillover effect are summed up as the total spillover effect.

### 3.3. Variables and Data Description

#### 3.3.1. Variables for Green Innovation Efficiency.

According to the principles of comprehensiveness, scientific, and availability of data, the index system for evaluating the efficiency of industrial green technology innovation is constructed by referring to the relevant research of green technology innovation:

- (1) Inputs: including labor input (number of R&D employees), capital input (total investment in R&D), and resource input (total energy consumption), which represent the consumption degree of innovation activities on resources.
- (2) Desirable outputs: including new product sales revenue and patent applications, which, respectively, reflect the economic benefits, living standards, and output level of scientific research achievements of each region.
- (3) Undesirable outputs: industrial pollution is the main source of environmental pollution, so the undesired output variable adopts the industrial wastewater discharge, industrial smoke (dust), and industrial sulfur dioxide discharge of each city in the YREB and uses the entropy method to calculate an environmental pollution index, which is used to explain the comprehensive impact of innovation activities on the ecological environment.

The input-output index system of green innovation efficiency in YREB is constructed in Table 1.

#### 3.3.2. Influential Factors on Green Innovation Efficiency.

There are many driving factors for the coordinated development of green innovation system. These factors will affect the development level of the subsystem invested in the regional green innovation subsystem and then affect the coordinated development level of the regional green innovation system. However, these factors cannot be used as the direct investment of each subsystem, and these factors are often not directly measured [47]. Therefore, in order to systematically and comprehensively study the driving factors of regional green innovation system, the factors that affect the efficiency of green innovation in the YREB are

TABLE 1: Evaluation indicator system of green innovation efficiency in YREB.

| Type                | Indicator        | Description   |
|---------------------|------------------|---|
| Inputs              | Labor            | Number of R&D employees (10,000 people)                         |
|                     | Capital          | Total investment in R&D (Billion yuan)                          |
|                     | Energy           | Total energy consumption (tons of standard coal)                |
| Desirable outputs   | Economic         | New product sales revenue (billion yuan)                        |
|                     | Technology       | Number of patent applications (billion)                         |
| Undesirable outputs | Industrial waste | Exhaust emissions, wastewater discharge, and solid waste (tons) |

summarized as direct and indirect factors, including environmental regulation and industrial structure, and the indirect factors include the level of economic development, the strength of government support, and the level of opening to the outside world [48]. Consider that the YREB, as a strategic region of our country, has different responsibilities and requirements in its upper, middle, and lower reaches. Based on the previous study, the following five factors are used to examine the impact on the green innovation efficiency:

- (1) Economic development (ED): green innovation has a higher threshold than traditional innovation. A higher level of economic development is conducive to the improvement of environmental protection needs and environmental human capital of residents and provides the necessary material basis and social environment for promoting the green innovation achievements. The exhibition has green incentive effect and cumulative effect of innovation ability, and it can promote the promotion of green innovation ability. The YREB is a national key construction inland river economic belt with global influence. Its economic development speed is at the national leading level, which should promote the ability of green innovation and enhance the competitiveness of regional green innovation development.
- (2) Environmental regulation (ER): Porter believes that environmental regulation can drive green innovation, which is the famous ‘‘Porter Hypothesis’’ [49]. Porter believes that appropriate environmental regulations can stimulate enterprises to increase investment in technology research and development, promote green innovation, and achieve a win-win situation of technological progress and environmental protection. Since the ‘‘Porter Hypothesis’’ was put forward, a large number of empirical research results show that environmental regulation is one of the important driving factors of green innovation [50–52]. Under the restriction of environmental regulation, the innovation subject in the region

should increase the investment in technological innovation, promote the improvement of the development level of technological innovation system and then reduce the expected output of the cost of achievement transformation subsystem, and promote the coordinated development of regional green innovation system. Therefore, environmental regulation is also an important driving factor for the coordinated development of regional green innovation system.

- (3) Government support (GS): technology driving factors are considered to be the fundamental cause of green innovation, and the improvement of technology capability triggers green innovation. This paper chooses government support for technology innovation to represent technology driving factors. The government's financial expenditure on science and technology improves the national green innovation ability and promotes economic growth. The government's support for scientific and technological innovation has created a good external environment for regional green innovation. To a large extent, the government's financial support also reflects the strength of the government's policy support. This paper chooses government funds from R&D funds as the indicator of government support for technological innovation. To a large extent, the government's financial support also reflects the strength of the government's policy support. Choose government funds from R&D funds as the indicator of government support for green innovation.
- (4) Foreign direct investment (FDI): the degree of market opening reflects the degree of exchange between a region and other regions in the fields of economy, science, and technology. The impact of market openness on green innovation is still controversial in academia. One of the most famous hypotheses is the "pollution shelter" hypothesis [53]. According to the "pollution shelter" hypothesis, companies in developed countries will transfer their pollution intensive industries to developing countries with relatively low regulation so that developing countries will become "pollution shelter paradise" and bear more environmental pollution [54]. But another hypothesis, pollution halo hypothesis, holds that market opening can reduce environmental pollution [55]. Through the spillover effect of foreign investment, developing countries bring advanced foreign green technologies, which can significantly improve the level of regional technological innovation and the level of regional unexpected output, thus promoting the coordinated development of regional green innovation system [56, 57]. The YREB covers the three major economic zones of the East, the middle, and the West. The introduction of foreign investment may promote the local technological progress, and at the same time, there will be

competition for foreign investment, which will make the surrounding cities backward in production capacity.

- (5) Industrial structure (IS): optimizing the internal allocation of the industry is conducive to stimulating the vitality of industrial innovation and enhancing the capacity of industrial green technology innovation. With the gradual upgrading of industrial structure, the secondary industry with strong pollution production capacity has transformed into a clean and low-carbon service industry, and the secondary and tertiary industries have accelerated the pace of integrated development [58, 59]. The close connection is promoted between green technology R&D services and industrial green transformation and enhanced the technological innovation ability with industrial characteristics. The YREB actively promotes the optimization and upgrading of industrial structure, promotes the integrated development of urban productive service industry and manufacturing industry, and requires enterprises to strengthen the research and development of green production technology to meet the technical requirements of industrial structure upgrading, and low-end production capacity may be forced to move to surrounding areas.

The influencing factors of green innovation efficiency in YREB is constructed in Table 2.

## 4. Empirical Analysis

*4.1. Green Innovation Efficiency of YREB.* Considering that there will be a certain time lag when green innovation input is converted into output, using other research results for reference, the input-output time lag is set as 1 year [60]; that is, the time interval of input index is set as 2008–2017, and the output index is set as 2008–2017. All the data were directly derived from the China Statistical Yearbook (2008–2017), the China Energy Statistical Yearbook (2008–2017), and the China Statistical Yearbook (2008–2017), the carbon dioxide emissions were estimated using the method provided by the Intergovernmental Panel on Climate Change [61]. Descriptive statistics of related variables are shown in Table 3. It can be preliminarily judged that the green innovation efficiency of 11 provinces and cities may also be significantly different, and further empirical analysis will be carried out in the future.

This paper relies on Max DEA PRO 8.0 software by using super-SBM model to measure the green innovation efficiency of 11 provinces in YREB from 2008 to 2017. The results are summarized in Table 4.

From 2008 to 2017, the overall green innovation efficiency of YREB was relatively stable. From 2008 to 2010, there was a slight downward trend. It increased significantly in 2013 and decreased slightly in 2013–2017. There are significant regional differences in green innovation efficiency level and time evolution trend in the upper, middle, and lower reaches of the YREB. During the research period,

TABLE 2: The influencing factors of green innovation efficiency in YREB.

| Variable                        | Description   |
|---------------------------------|---|
| Economic development (ED)       | GDP per capita  |
| Environmental regulation (ER)   | The ratio of total investment in industrial pollutants to GDP                     |
| Government support (GS)         | R&D funds   |
| Foreign direct investment (FDI) | The proportion of foreign investment as a percentage of the regional GDP          |
| Industrial structure (IS)       | Proportion of total output value of tertiary industry to total GDP in each region |

TABLE 3: Descriptive statistics of green innovation efficiency in YREB.

| Index  | Minimum | Maximum | Mean    | Standard deviation |
|--|---------|---------|---------|--------------------|
| Number of R&D employees (10,000 people)          | 12656   | 466735  | 111834  | 134245             |
| Total investment in R&D (Billion yuan)           | 324986  | 4365780 | 4326382 | 6023576            |
| Total energy consumption (tons of standard coal) | 4658    | 30480   | 13762   | 6187               |
| New product sales revenue (billion yuan)         | 3795210 | 7456754 | 7134578 | 7238568            |
| Number of patent applications (billion)          | 1785    | 125784  | 32650   | 33468              |
| Exhaust emissions (10,000 tons)                  | 8730    | 64390   | 23561   | 13652              |
| Wastewater discharge (10,000 tons)               | 14370   | 455321  | 138542  | 100654             |
| Solid waste (10,000 tons)                        | 1450    | 17890   | 8974    | 4376               |

TABLE 4: Green innovation efficiency of YREB in 2008–2017.

| Region    | Year  |       |       |       |       |       |       |       |       |       |
|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|           | 2008  | 2009  | 2010  | 2011  | 2012  | 2013  | 2014  | 2015  | 2016  | 2017  |
| Jiangsu   | 1.032 | 1.053 | 1.047 | 1.092 | 1.085 | 1.066 | 1.023 | 1.026 | 1.024 | 1.031 |
| Shanghai  | 1.042 | 1.035 | 1.037 | 1.032 | 1.013 | 1.021 | 1.032 | 1.027 | 1.034 | 1.089 |
| Zhejiang  | 0.765 | 0.763 | 0.768 | 0.827 | 0.845 | 0.976 | 0.853 | 0.812 | 0.743 | 0.751 |
| Anhui     | 0.652 | 0.654 | 0.667 | 0.706 | 0.71  | 0.733 | 0.724 | 0.697 | 0.681 | 0.678 |
| Jiangxi   | 0.622 | 0.623 | 0.631 | 0.673 | 0.668 | 0.677 | 0.682 | 0.653 | 0.646 | 0.643 |
| Hubei     | 0.649 | 0.661 | 0.667 | 0.711 | 0.733 | 0.742 | 0.744 | 0.725 | 0.711 | 0.698 |
| Hunan     | 0.68  | 0.671 | 0.68  | 0.731 | 0.745 | 0.947 | 1.012 | 0.98  | 1.092 | 1.001 |
| Chongqing | 0.625 | 0.63  | 0.639 | 0.688 | 0.705 | 0.698 | 0.707 | 0.715 | 0.689 | 0.691 |
| Sichuan   | 0.625 | 0.633 | 0.643 | 0.707 | 0.722 | 0.703 | 0.702 | 0.686 | 0.664 | 0.653 |
| Guizhou   | 0.596 | 0.591 | 0.59  | 0.635 | 0.632 | 0.636 | 0.634 | 0.638 | 0.645 | 0.63  |
| Yunnan    | 0.619 | 0.601 | 0.602 | 0.628 | 0.631 | 0.632 | 0.636 | 0.63  | 0.615 | 0.613 |
| Average   | 0.719 | 0.720 | 0.725 | 0.766 | 0.772 | 0.803 | 0.795 | 0.781 | 0.777 | 0.771 |

the green innovation efficiency of the middle and lower reaches of the province showed a trend of increasing first and then decreasing slightly, while that of the upper and lower reaches showed a trend of decreasing first and then increasing and then decreasing, but the fluctuation range was small. Since the promulgation of several opinions of the State Council on promoting the rise of the central region in 2005, the industrial undertaking policies to promote the rise of the central and western regions have promoted the inflow of a large number of capital and labor factors, which has led to the economic growth of the central and western regions. However, the industrial undertaking has brought economic benefits as well as unexpected output, making the green innovation efficiency of the middle and upper reaches of provinces at it is low and declining, and only in recent years, does it show an upward trend. In 2014, the policy of building the YREB into a leading demonstration zone of ecological civilization was issued. Since then, the state and local governments of the YREB have successively issued relevant

TABLE 5: Global Moran's I Index of green innovation efficiency.

|      | Moran's I | z     |
|------|-----------|-------|
| 2008 | 0.540***  | 3.544 |
| 2009 | 0.510***  | 3.415 |
| 2010 | 0.514***  | 3.432 |
| 2011 | 0.533***  | 3.476 |
| 2012 | 0.487***  | 3.258 |
| 2013 | 0.313**   | 2.022 |
| 2014 | 0.252**   | 1.756 |
| 2015 | 0.233**   | 1.622 |
| 2016 | 0.226*    | 1.634 |
| 2017 | 0.209*    | 1.567 |

Note: \*\*\*, \*\*, and \* represent the significance level at 1%, 5%, and 10%.

policies, and the construction of ecological civilization in the YREB has achieved initial results.

#### 4.2. Spatial Autocorrelation Analysis

4.2.1. *Global Spatial Autocorrelation.* This paper uses MATLAB to calculate the Global Moran's I Index of logistics green innovation efficiency in YREB. Table 5 presents the results.

The results show that the Global Moran's I Index of the green innovation efficiency passed the test at 10% significance level in 2008–2017, indicating that the green innovation efficiency displays a positive spatial correlation, it is not distributed randomly, the index is basically between 0.3 and 0.6, and it indicates that the green innovation efficiency shows a weak agglomeration state. In the main, the Global Moran's I Index moves upward along a wave-like curve, this may be related to the macroeconomic environment in which the Chinese economy enters the “new normal,” and the industrial structure is transformed and upgraded.

The high and low efficiency neighboring provinces show a spatial cluster. The higher green innovation efficiency regions were adjacent, and the regions with lower green innovation efficiency were close to each other.

4.2.2. *Spatial Effect of Green Innovation Efficiency.* Because the data used in this paper are panel data, it is necessary to determine whether the fixed effect model or the random effect model should be used before regression analysis of the model. The Hausman test was carried out for SLM and SEM by MATLAB, and the test results are shown in Table 6.

According to Table 6, both SLM and SEM passed the Hausman test at 5% significance level, so the panel model with fixed effect was selected for regression analysis.

According to the test of spatial correlation, Table 7 shows that the test of LM\_lag is greater than LM\_error, so the estimation method of spatial lag model is studied and analyzed.

The general OLS regression coefficient is smaller than the spatial Durbin model (SDM), which shows that OLS regression ignores the spatial interaction between independent variables and dependent variables and overestimates the influence of related variables. From the regression results of the SDM model, the log-L and  $R^2$  of the spatiotemporal fixed model are obvious, and it is larger than the fixed time model and the fixed space model, so the double fixed model has the best estimation results. Table 8 is an analysis of the influencing factors of green innovation efficiency based on the empirical results of the time-space fixed model. The results are shown in Table 8.

- (1) Economic development (ED) has a significant positive role in promoting the green innovation efficiency of the YREB. For every 1% increase in GDP per capita, the efficiency of green innovation will increase by an average of 0.3487%. It shows that economic growth will improve the green innovation efficiency. Economically developed regions in the YREB, on the one hand, will pay more attention to the development of environmental quality; on the other hand, the R&D investment in the field of green innovation will increase, and the investment

TABLE 6: Hausman test results.

| Test summary | Hausman test-statistic | Variance |
|--------------|------------------------|----------|
| SLM          | 29.631***              | 15       |
| SEM          | 231.586***             | 29       |

Note: \*\*\*, \*\*, and \* represent the significance level at 1%, 5%, and 10%.

TABLE 7: Spatial correlation test results.

| Spatial dependence test | LM_lag    | Robust LM_lag | LM_error  | Robust LM_error |
|-------------------------|-----------|---------------|-----------|-----------------|
| □                       | 68.237*** | 19.632***     | 50.792*** | 10.011***       |

Note: \*\*\*, \*\*, and \* represent the significance level at 1%, 5%, and 10%.

TABLE 8: Estimation and test results based on spatial Durbin model (SDM) for the driving factor.

|         | OLS               | TF                  | SF                  | STF                 |
|---------|-------------------|---------------------|---------------------|---------------------|
| LnED    | 0.1813***<br>7.17 | 0.1255***<br>4.23   | 0.2536***<br>5.36   | 0.3487***<br>6.91   |
| LnIS    | 0.0521***<br>0.82 | 0.074***<br>1.32    | 0.0867***<br>1.43   | 0.1356***<br>0.38   |
| LnFDI   | 0.023***<br>0.639 | 0.017<br>2.35       | 1.551<br>3.75       | 0.027<br>4.11       |
| LnGS    | 0.003***<br>0.38  | 0.003***<br>0.13    | 0.030**<br>2.45     | 0.029***<br>2.99    |
| LnER    | 0.002***<br>1.17  | 0.003***<br>2.04    | 0.019***<br>2.91    | 0.005***<br>0.67    |
| W*LnED  |                   | -0.1675***<br>-2.34 | -0.2036***<br>-3.12 | -0.1593***<br>-4.02 |
| W*LnIS  |                   | 0.074***<br>1.32    | 0.0867***<br>1.22   | 0.1356***<br>0.46   |
| W*LnFDI |                   | 0.017*<br>2.35      | 1.551<br>4.00       | 0.027<br>2.06       |
| W*LnGS  |                   | 0.001***<br>0.13    | 0.030**<br>2.40     | 0.029***<br>2.99    |
| W*LnER  |                   | 0.004***<br>-0.44   | 0.021***<br>-0.22   | 0.003***<br>-0.44   |
| $\rho$  | 0.434***<br>7.57  | -0.464***<br>7.53   | -0.003***<br>4.76   | -0.117***<br>0.96   |
| $R^2$   |                   | 0.642               | 0.643               | 0.709               |
| log-L   |                   | 568.895             | 683.685             | 709.185             |

\*\*\*, \*\*, and \* represent the significance level at 1%, 5%, and 10%.

subsidies and production subsidies for products and services will be greater. The coefficient of the spatial lag term of economic growth is  $-0.1593$ , and through the 1% significance test, it shows that the economic growth of the neighboring areas in the YREB has negative spatial spillover effects to the green innovation efficiency of the region. This is because the economic development of the neighboring areas will have a certain siphon effect on the relevant innovation elements of the region, which is not conducive to the improvement of the green innovation efficiency of the region.

- (2) Industrial structure (IS) has a significant positive role at the level of 5%. In the TF model, the coefficient is also significantly positive. This is mainly because in the adjustment and upgrading of industrial structure in the YREB, those high energy consumptions and high pollution situation have improved. However, in the process of promoting the industrial structure to achieve a high degree and rationalization in a certain region in the YREB, it may cause the imitation of neighboring regions and promote the coordinated development of regional green systems in different regions.
- (3) Foreign direct investment (FDI) has no significant effect on the green innovation efficiency of the YREB. At the same time, the corresponding spatial lag has not passed the significance test, which means that when the YREB regions introduce FDI, they neither promote the efficiency of local green innovation nor bring spillover effects to the efficiency of green innovation in the surrounding areas. The reason may be that FDI does not really consider regional environmental technology innovation in most provinces or even occupies the provincial R&D innovation of the YREB and inhibits the technological innovation ability, and the green innovation efficiency of introducing foreign investment is not ideal. This also means that the purpose of most FDI entry is to pursue low cost and tax advantages. It does not really consider environmental technology innovation, and the quality of investment still needs to be further improved.
- (4) Government support (GS) has a significant positive role in promoting the green innovation efficiency of the YREB. The government's support for green innovation activities can improve the development level of scientific and technological research and development subsystem, so as to reduce the unexpected output in the process of achievement transformation, increase the expected output, and promote the coordinated development of green innovation system. The government should continue to increase its support for green innovation in the YREB, especially in the less developed areas such as the central and western regions. By promoting the development of technological innovation, we can develop more technologies that are beneficial to the ecological environment and promote the coordinated development of green innovation system.
- (5) Environmental regulation (ER) has significant positive effect on the green innovation efficiency of the YREB. For every 1% increase in the level of environmental regulation, the green innovation efficiency will increase by an average of 0.005%, which means that the more stringent the environmental regulation is, the stronger the environmental pollution cost constraints enterprises bear, so that they have the motivation to pay attention to the

production of clean, ecological, and recycling, and the enterprises that take the lead in technological innovation have the first mover advantage in pollution control. It is helpful for enterprises to seize market share and gain competitive advantage, and it is also helpful for enterprises to improve their green innovation performance. The coefficient of the spatial lag term of environmental regulation is  $-0.003$ , and through the 1% significance test, it shows that the environmental regulation of the neighboring areas has negative spatial spillover benefits to the green innovation efficiency of the region. This may be due to the deterrence effect of environmental regulations on enterprises, which forces enterprises to increase investment in environmental governance, so that the corresponding low-tech pollution links are transferred to other areas with relatively low environmental standards, leading to the "pollution shelter" effect and inhibiting green innovation.

#### 4.2.3. Spatial Spillover Effects of Green Innovation Efficiency.

Based on the SDM model, this paper analyzes the direct, indirect, and total effects of various influencing factors, among which the total effect represents the average impact of influencing factors on green innovation, while the direct effect and indirect effect represent the decomposition of the total effect, which, respectively, represents the impact of influencing factors on the region and adjacent regions. Table 9 are the results.

The indirect effect of the level of economic development is negative, which shows that, in general, the level of economic development is not conducive to the spatial spillover of green innovation efficiency in the period under investigation, especially in the provinces with the higher level of economic development, the lower the spatial spillover effects of green innovation efficiency, which highlights that the more developed provinces pay more attention to "protect" the efficiency of green innovation in their own provinces in the YREB. Under the requirements of national green development, the economically developed regions in the YREB make use of their own advantages in capital, and the economically underdeveloped regions are eager to improve the local economic level and transfer some high energy consumption, high pollution, and high emission industries to the economically underdeveloped provinces, which to some extent causes the green innovation efficiency to show negative spatial spillover effects.

The total effect, direct effect, and indirect effect of industrial structure are all positive. The optimization of industrial structure is conducive to the transformation of development mode, reduction of energy consumption, and environmental pollution, so as to improve the efficiency of green innovation.

FDI in this region has no significant impact on the green innovation efficiency of surrounding areas in the YREB. It shows that there are corresponding regional technical barriers in the process of promoting the

TABLE 9: Spatial spillover effects of green innovation efficiency.

|       | Direct effect    | Indirect effect   | Total effect     |
|-------|------------------|-------------------|------------------|
| LnED  | 0.244***<br>5.20 | -0.232***<br>4.48 | 0.012***<br>8.82 |
| LnIS  | 0.038***<br>1.29 | 0.047***<br>0.89  | 0.085***<br>1.27 |
| LnFDI | 0.162*<br>1.45   | 0.414<br>3.68     | 0.576<br>2.87    |
| LnGS  | 0.025***<br>2.56 | 0.034<br>3.28     | 0.059**<br>2.62  |
| LnER  | 0.381***<br>1.41 | -0.463**<br>1.86  | -0.082*<br>1.12  |

\*\*\*, \*\*, and \* represent the significance level at 1%, 5%, and 10%.

efficiency of green innovation in all regions of YREB. On the one hand, they are unwilling to cooperate and share the technical innovation; on the other hand, they are easy to transfer the cost of environmental pollution to the surrounding areas.

The direct effect regression coefficient is positive under the significance of 5%, and the indirect effect regression coefficient fails to pass the significance test. Government support is a strong backing to enhance the capacity of green innovation, especially along with the transformation and upgrading driven by green innovation in the YREB, and the government has given strong support in building basic innovation platform and increasing investment in innovation and R&D. The guiding effect of green innovation policy is significant, laying a solid foundation for the high-quality development of the YREB.

The direct effect of environmental regulation on the spatial spillover of green innovation efficiency is positive and passes the significance test; the indirect effect and the total effect on the productivity of green innovation pass the significance test of 5% but has a negative impact on the change in green innovation efficiency. This shows that, on the one hand, environmental regulation has a positive role in promoting the efficiency of green innovation in our province, but at present, it has not fundamentally changed the level of green development in China, so it cannot significantly improve the efficiency of green innovation in China. On the other hand, when China's green innovation capacity is insufficient, the imbalance of the intensity of interprovincial environmental regulation is likely to lead the environmental pollution industry in the provinces with high intensity of regulation to enter the provinces with low intensity of regulation in the YREB.

## 5. Conclusions and Discussion

*5.1. Conclusions.* This study used super-SBM model to consider undesirable outputs, measuring the green innovation efficiency in YREB from 2008 to 2017. Since green innovation efficiency has spatial spillover effects, therefore, a spatial econometric model SDM model is applied to analyze the influencing factors of green innovation efficiency. The green innovation efficiency empirical results indicate the green innovation efficiency is developing slowly, and the

green innovation of the eastern part of YREB is significantly better than that of the lower reaches in the west. From the spatial autocorrelation result, it shows that there is a significant spatial autocorrelation of green innovation efficiency in YREB regions. From the spatial econometric of SDM analysis, indicating that the level of economic development, foreign direct investment to the outside world and environmental pollution control has positive effects on the green economic efficiency of the YREB, while the proportion of the secondary industry has negative effects. The green economic efficiency of the YREB has a significant spatial correlation. The provinces with high level of economic development and environmental pollution control have a significant positive role in promoting the green economic efficiency of the neighboring provinces. The provinces with high proportion of the secondary industry and high government support have a negative inhibitory effect on the green economic efficiency of the neighboring provinces.

*5.2. Discussion.* According to the empirical results, this study put forward proposals to enhance green innovation efficiency.

First, the optimization and upgrading of industrial structure is promoted. Industrial structure has a significant negative inhibitory effect on the green innovation efficiency of the YREB, so it is necessary to speed up the pace of industrial structure adjustment and new industrialization. We will bring superiority into full play of industry and intelligence intensity in the YREB, vigorously implement innovation-driven development strategy, add to new momentum of reform, innovation, and development, subtract from the elimination of backward production capacity, and accelerate industrial transformation and upgrading. We will build a manufacturing innovation system, improve the ability to develop key systems and equipment, and foster and expand high technology industries, emerging sectors of strategic importance, equipment manufacturing, and other industries. We will optimize the layout of strategic emerging industries, accelerate the construction of regional characteristic industrial bases, give free rein to radiation driving and leading demonstration, and form a national strategic emerging industry development highland.

Second, the quality of opening up is improved. The technology spillover effects of FDI in the YREB are more than the environmental pollution effect. The introduction of FDI can improve green economic development level in the YREB, but improvement effect is not significant, so we should further improve the quality of opening up. We should further promote the improvement of the negative list of market access in the YREB, improve the project access mechanism, promote the formation of an institutionalized, standardized, green, transparent, and procedural system for foreign capital introduction system, vigorously introduce new green technologies and industries, attract environmentally friendly enterprises to settle down, give full play to the technology spillover effects of green foreign capital industries, and improve the green production of local

enterprises to build a green ecological industrial chain and improve the level of green development and ecological quality.

Finally, we will intensify efforts to prevent and control environmental pollution. Environmental pollution control in the YREB has a positive role in improving green innovation efficiency, but it has not produced significant results, so the efforts of environmental pollution control need to be further strengthened. We should strengthen the joint prevention and control of environmental pollution; establish and improve the emergency response mechanism for cross department, cross region, and cross basin environmental emergencies; strictly control industrial pollution; dispose of urban sewage and garbage; control agricultural nonpoint source pollution; prevent ship and air pollution; strengthen the collaborative protection of ecological environment; establish a negative list management system; strengthen daily monitoring and supervision; and strictly implement the ecological environment. The system of responsibility investigation for environmental damage should be improved, the proportion of resource utilization rate, environmental pollution prevention and control, and quality evaluation system of economic and ecological development should be increased, and the performance evaluation system reflecting the requirements of ecological civilization should be improved.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Disclosure

Hangyuan Guo is the co-first author.

## Conflicts of Interest

The authors declare that there are no conflicts of interest.

## Acknowledgments

This paper was supported by Wonkwang University in 2020.

## References

- [1] Y. Yuan, R. Luo, and Y. Li, "Analysis of the development level and spatial evolution characteristics of ecological civilization in the Yangtze river economic belt," *Statistics and Decision*, vol. 1, no. 20, pp. 98–101, 2016.
- [2] L. W. Lu, D. Y. Song, and X. F. Li, "Research on green efficiency of urban development in the Yangtze river economic belt," *China Population, Resources and Environment*, vol. 26, no. 6, pp. 35–42, 2016.
- [3] Q. Luo, C. Miao, L. Sun, X. Meng, and M. Duan, "Efficiency evaluation of green technology innovation of China's strategic emerging industries: an empirical analysis based on Malmquist-data envelopment analysis index," *Journal of Cleaner Production*, vol. 238, Article ID 117782, 2019.
- [4] C. Ghisetti, S. Mancinelli, M. Mazzanti, and M. Zoli, "Financial barriers and environmental innovations: evidence from EU manufacturing firms," *Climate Policy*, vol. 17, no. sup1, pp. S131–S147, 2017.
- [5] D. Li, M. Zheng, C. Cao, X. Chen, S. Ren, and M. Huang, "The impact of legitimacy pressure and corporate profitability on green innovation: evidence from China top 100," *Journal of Cleaner Production*, vol. 141, pp. 41–49, 2017.
- [6] M. Cao, J. Ukko, and T. Rantala, "Sustainability as a driver of green innovation investment and exploitation," *Journal of Cleaner Production*, vol. 179, pp. 631–641, 2018.
- [7] T. Bernauer, S. Engel, and D. Kammerer, "Explaining green innovation: ten years after porter's win-win proposition: how to study the effects of regulation on corporate environmental innovation," *Politische Vierteljahresschrift*, vol. 39, pp. 323–341, 2007.
- [8] C. Fussier and P. James, *Book Review: Driving Eco-Innovation: A Breakthrough Discipline for Innovation and Sustainability*, [CrossRef], p. 297, Pitman Pub, Wetherby, UK Pitman Pub, 1996.
- [9] R. Kemp and A. Arundel, Survey indicators for environmental innovation. IDEA (indicators and data for European analysis). Sub-project 2.2, Indicators on the importance of environmental goals. 26, 1998.
- [10] L. Ahlvik, P. Ekholm, K. Hyytiäinen, and H. Pitkänen, "An economic-ecological model to evaluate impacts of nutrient abatement in the Baltic Sea," *Environmental Modelling & Software*, vol. 55, pp. 164–175, 2014.
- [11] T. Heffels, R. McKenna, and W. Fichtner, "An ecological and economic assessment of absorption-enhanced-reforming (AER) biomass gasification," *Energy Conversion and Management*, vol. 77, pp. 535–544, 2014.
- [12] Y. Chen, C. Jayaprakash, and E. Irwin, "Threshold management in a coupled economic-ecological system," *Journal of Environmental Economics and Management*, vol. 64, no. 3, pp. 442–455, 2012.
- [13] Z. J. Feng and W. Chen, "Sources of technology and the total factor productivity growth of R & D innovation: based on China regional big medium-sized industrial enterprises," *Science of Science and Management of S & T*, vol. 34, no. 3, pp. 33–41, 2013.
- [14] Y. Ren, C. K. Niu, T. Niu, and X. L. Yao, "Research on the green innovation efficiency model and empirical analysis," *Management World*, vol. 7, pp. 176–177, 2014.
- [15] Q. Yin and Y. Chen, "Study on the regional differences and causes of green innovation efficiency in China," *Jiangsu Social Sciences*, vol. 18, no. 2, pp. 64–69, 2016, [CrossRef].
- [16] T. Charoenrat and C. Harvie, "The efficiency of SMEs in Thai manufacturing: a stochastic frontier analysis," *Economic Modelling*, vol. 43, pp. 372–393, 2014.
- [17] H. Li, J. Zhang, C. Wang, Y. Wang, and V. Coffey, "An evaluation of the impact of environmental regulation on the efficiency of technology innovation using the combined DEA model: a case study of Xi'an, China," *Sustainable Cities and Society*, vol. 42, pp. 355–369, 2018.
- [18] H. W. Lampe and D. Hilgers, "Trajectories of efficiency measurement: a bibliometric analysis of DEA and SFA," *European Journal of Operational Research*, vol. 240, no. 1, pp. 1–21, 2015.
- [19] C. Miao, D. Fang, L. Sun, Q. Luo, and Q. Yu, "Driving effect of technology innovation on energy utilization efficiency in strategic emerging industries," *Journal of Cleaner Production*, vol. 170, pp. 1177–1184, 2018.
- [20] L. Xiao, J. Gao, and S. Liu, "The change trend of green technology innovation efficiency in China based on spatial



- gradient—empirical analysis of provincial panel data,” *Soft Science*, vol. 31, pp. 63–68, 2017.
- [21] E. G. Carayannis, E. Grigoroudis, and Y. Goletsis, “A multilevel and multistage efficiency evaluation of innovation systems: a multiobjective DEA approach,” *Expert Systems with Applications*, vol. 62, pp. 63–80, 2016.
- [22] P. Samut and R. Cafri, “Analysis of the efficiency determinants of health systems in OECD countries by DEA and panel tobit,” *Social Indicators Research*, vol. 129, no. 1, pp. 113–132, 2016.
- [23] C. Lafarga and J. Balderrama, “Efficiency of Mexico’s regional innovation systems: an evaluation applying data envelopment analysis (DEA),” *African Journal of Science Technology Innovation & Development*, vol. 7, pp. 36–44, 2015.
- [24] Y. Ren and C. Wang, “Research on the regional difference and spatial effect of green innovation efficiency of industrial enterprises in China,” *Revista Ibérica de Sistemase Tecnologias de Informação*, vol. 10, pp. 373–384, 2016.
- [25] J.-L. Du, Y. Liu, and W.-X. Diao, “Assessing regional differences in green innovation efficiency of industrial enterprises in China,” *International Journal of Environmental Research and Public Health*, vol. 16, no. 6, p. 940, 2019.
- [26] M. Buesa, J. Heijs, M. Martínez Pellitero, and T. Baumert, “Regional systems of innovation and the knowledge production function: the Spanish case,” *Technovation*, vol. 26, no. 4, pp. 463–472, 2006.
- [27] R. Färe, S. Grosskopf, and G. Whittaker, “Network DEA II,” in *Data Envelopment Analysis. International Series in Operations Research & Management Science*, W. Cook and J. Zhu, Eds., vol. 208, pp. 307–327, Springer, Berlin, Germany, 2014.
- [28] J. Guan and K. Zuo, “A cross-country comparison of innovation efficiency,” *Scientometrics*, vol. 100, no. 2, pp. 541–575, 2014.
- [29] X. Yu, Z. Li, R. Chi, and M. Shi, “Technological innovation efficiency of different regions in China: status quo and causes,” *Studies In Science of Science*, vol. 23, pp. 258–264, 2005.
- [30] H. Wang, S. Wang, Z. Miao, and X. Li, “Heterogeneity threshold effect of R & D investment on green innovation efficiency based on Chinese high-tech industries,” *Science Research Management*, vol. 37, pp. 63–71, 2016.
- [31] S. Yu, X. Li, and Z. Peng, “Environmental regulation model and green innovation efficiency of the Yangtze river economic belt,” *Jianghai Academic Journal*, vol. 3, pp. 209–214, 2017.
- [32] S. Yang, T. Wu, and Z. Li, “Study on the spatial-temporal differences and influencing factors of green innovation efficiency in the Yangtze river economic belt,” *MACROECONOMICS*, vol. 6, pp. 107–132, 2018.
- [33] X. Gong, M. Li, and H. Zhang, “Has OFDI promoted the industrial enterprises’ green innovation efficiency in China—evidence based on agglomeration economic effect,” *Journal of International Trade*, vol. 11, pp. 127–137, 2017.
- [34] W. D. Cook and L. M. Seiford, “Data envelopment analysis (DEA)—Thirty years on,” *European Journal of Operational Research*, vol. 192, no. 1, pp. 1–17, 2009.
- [35] H. Li, K. Fang, W. Yang, D. Wang, and X. Hong, “Regional environmental efficiency evaluation in China: analysis based on the super-SBM model with undesirable outputs,” *Mathematical and Computer Modelling*, vol. 58, no. 5–6, pp. 1018–1031, 2013.
- [36] J. Zhang, W. Zeng, J. Wang, F. Yang, and H. Jiang, “Regional low-carbon economy efficiency in China: analysis based on the super-SBM model with CO<sub>2</sub> emissions,” *Journal of Cleaner Production*, vol. 163, pp. 202–211, 2017.
- [37] K. Tone, “A slacks-based measure of efficiency in data envelopment analysis,” *European Journal of Operational Research*, vol. 130, no. 3, pp. 498–509, 2001.
- [38] K. Tone, “A slacks-based measure of super-efficiency in data envelopment analysis,” *European Journal of Operational Research*, vol. 143, no. 1, pp. 32–41, 2002.
- [39] F. Jin and L. Lee, “On the bootstrap for Moran’s I test for spatial dependence,” *Journal of Economics*, vol. 184, pp. 295–314, 2015.
- [40] S. J. Rey, “Spatial empirics for economic growth and convergence,” *Geographical Analysis*, vol. 33, no. 3, pp. 195–214, 2001.
- [41] Y. Xiong, D. Bingham, W. J. Braun, and X. J. Hu, “Moran’s I statistic-based nonparametric test with spatio-temporal observations,” *Journal of Nonparametric Statistics*, vol. 31, no. 1, pp. 244–267, 2019.
- [42] T. Zhang and G. Lin, “On Moran’s I coefficient under heterogeneity,” *Computational Statistics & Data Analysis*, vol. 95, pp. 83–94, 2016.
- [43] H. D. Liu, “The inside, outside, and space spillover effects of regional innovation,” *Science Research Management*, vol. 34, no. 1, pp. 28–36, 2013.
- [44] J. P. Elhorst, “Dynamic spatial panels: models, methods, and inferences,” *Journal of Geographical Systems*, vol. 14, no. 1, pp. 5–28, 2012.
- [45] J. P. LeSage and R. K. Pace, *Introduction to Spatial Econometrics*, CRC Press, Boca Raton, FL, USA, 2009.
- [46] K. R. Zuo and J. C. Gong, “Exploring the change and influence factors of R & D efficiency at province-level of China,” *Science of Science and Management of S & T*, vol. 37, no. 4, pp. 79–88, 2016.
- [47] H. Chen, H. Lin, and W. Zou, “Research on the regional differences and influencing factors of the innovation efficiency of China’s high-tech industries: based on a shared inputs two-stage network DEA,” *Sustainability*, vol. 12, no. 8, p. 3284, 2020.
- [48] K. Rennings, “Redefining innovation-eco-innovation research and the contribution from ecological economics,” *Ecological Economics*, vol. 32, no. 2, pp. 319–332, 2000.
- [49] M. E. Porter and C. Van Der Linde, “Green and competitive: ending the stalemate,” *Harvard Business Review*, vol. 73, no. 5, pp. 120–134, 1995.
- [50] J. Horbach, C. Rammer, and K. Rennings, “Determinants of eco-innovations by type of environmental impact—the role of regulatory push/pull, technology push and market pull,” *Ecological Economics*, vol. 78, pp. 112–122, 2012.
- [51] E. Kesidou and P. Demirel, “On the drivers of eco-innovations: empirical evidence from the UK,” *Research Policy*, vol. 41, no. 5, pp. 862–870, 2012.
- [52] K. Rennings and C. Rammer, “The impact of regulation-driven environmental innovation on innovation success and firm performance,” *Industry & Innovation*, vol. 18, no. 3, pp. 255–283, 2011.
- [53] B. R. Copeland and M. S. Taylor, “North-South trade and the environment,” *The Quarterly Journal of Economics*, vol. 109, no. 3, pp. 755–787, 1994.
- [54] J. X. Zhang, N. Cai, J. S. Mao, and C. Yang, “Independent innovation, technology introduction and green growth of industry in China: an empirical research based on industry heterogeneity,” *Studies in Science of Science*, vol. 33, no. 2, pp. 185–194, 2015.
- [55] M. H. Kim and N. Adilov, “The lesser of two evils: an empirical investigation of foreign direct investment-pollution

- tradeoff,” *Applied Economics*, vol. 44, no. 20, pp. 2597–2606, 2012.
- [56] K. X. Bi, C. J. Yang, and P. Huang, “Research on the impact of FDI on the green process innovation of Chinese manufacturing industries: an empirical analysis based on the panel data,” *China Soft Science*, vol. 20, no. 9, pp. 172–180, 2011.
- [57] J. Horbach, “Determinants of environmental innovation-new evidence from German panel data sources,” *Research Policy*, vol. 37, no. 1, pp. 163–173, 2008.
- [58] M. C. Cuerva, A. Triguero-Cano, and D. Córcoles, “Drivers of green and non-green innovation: empirical evidence in low-tech SMEs,” *Journal of Cleaner Production*, vol. 68, pp. 104–113, 2014.
- [59] L. Luo and S. Liang, “Study on the efficiency and regional disparity of green technology innovation in China’s industrial companies,” *Chinese Journal of Population Resources and Environment*, vol. 14, no. 4, pp. 262–270, 2017.
- [60] P. C. Zhu, D. H. Liu, and X. H. Huang, “An evaluation of science and technology innovation efficiency of cities from the dynamic perspective by taking 9 prefecture-level cities in Fujian Province as an example,” *Science Research Management*, vol. 38, no. 6, pp. 43–50, 2017.
- [61] R. Y. Long, H. Z. Ouyang, and H. Y. Guo, “Super-slack-based measuring data envelopment analysis on the spatial-temporal patterns of logistics ecological efficiency using global Malmquist index model,” *Environmental Technology & Innovation*, vol. 18, Article ID 100770, 2020.

## Research Article

# Optimization and Simulation of Controller Area Network Communication Model Based on Industrial Internet of Things Platform

Haifeng Lin <sup>1</sup> and Lin Du <sup>2</sup>

<sup>1</sup>College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, China

<sup>2</sup>School of Information Science and Engineering, Qilu Normal University, Jinan 250200, China

Correspondence should be addressed to Lin Du; [dul1028@163.com](mailto:dul1028@163.com)

Received 19 August 2020; Revised 27 September 2020; Accepted 30 September 2020; Published 26 October 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Haifeng Lin and Lin Du. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The Internet of Things (IoT) is an information carrier based on network communication, and it can make all general physical objects which can be independently addressed from an interconnected network. Controller Area Network (CAN) is a kind of serial communication network which can effectively support distributed control or real-time control. This paper proposes the design of an IoT-oriented universal CAN bus. In order to build a complete CAN communication simulation model, message sending and message receiving models are constructed, respectively. The simulation experiment proves that the IoT-oriented network communication model designed in this paper not only has low power consumption, but it can also solve the front-end compatibility problems caused by different communication protocols.

## 1. Introduction

Internet of Things technology has been listed as one of the emerging industries with strategic significance in the world and has developed rapidly. IoT is an object-oriented Internet. Its three major technologies, sensor technology, communication technology, and embedded system technology, correspond to its three attributes of perception, transmission, and control [1]. Among these three technologies, embedded system technology is the foundation, sensor technology is the function expansion, and the communication technology is the core connection technology of IoT [2]. The communication technology of IoT includes four types: wire transmission, wireless transmission, traditional Internet, and mobile air network. The IoT system does not need every device to be connected to the Internet; it can have one device as the control core, to which the other devices are connected through wireless communication technology and this device is connected to the Internet for communication [3]. This method can lower the

application cost of IoT. CAN bus is a serial communication network that can effectively support distributed control and real-time control and it has been widely applied in the field of automatic control with its high performance and reliability. The CAN protocol has multimaster control and it also has the functions of error detection, error notification, and error recovery. All units can detect error and the unit which detects the error will notify all other units immediately [4]. Besides, CAN bus has many connection nodes, and it can be connected to the bus with multiple units. Theoretically, there is no limit for the number of connectable units. Besides, it has a fast communication speed and a long communication distance. The fastest speed is 1 Mbps (with the distance less than 40 m) and the longest distance is 10 km (with the speed slower than 5 kbps) [5]. In field bus, CAN bus has obtained extensive support of computer chip makers, who have released microcontroller unit (MCU) chip with direct CAN interface one after another. With the development of CAN bus technology, it has already become one of field buses that has been widely applied in many fields,

including agriculture, robot, computer numerically controlled machine tool, medical apparatus and instruments, vehicles, vessels, and national defense [6].

The special contributions of this paper include the following:

- (i) It has firstly elaborated the significance and background of this paper, analyzed the research status at home and abroad of IoT and CAN bus, and expounded the related theoretical foundation.
- (ii) It has introduced the technological architecture and logical architecture of industrial equipment IoT platform, analyzed CAN real-time communication technology, and studied the network communication, socket mechanism, and breakpoint resume with reliable time sequence.
- (iii) It has proposed an IoT-oriented universal CAN bus and explored some problems to which attention should be paid in application. It has also proposed the design method for CAN bus interface circuit for its highly integrated communication controller.
- (iv) It has effectively simulated and constructed CAN communication, constructed the complete model structure, and conducted simulation and operation to check its effect. It has proven that the CAN communication design scheme of this paper is effective.

The remainder of this paper is organized as follows. Section 2 discusses related works, and IoT-oriented CAN communication design is outlined in Section 3. Experimental procedure is presented in Section 4. Section 5 shows the experimental simulation results, and Section 6 concludes the paper with a summary and proposed directions for future research.

## 2. Related Work

With the development of IoT, communication technology, and control technology, conventional industrial control field has been experiencing an unprecedented reform and the networked industrial control has expanded the development space for industrial control field and brought new development opportunities [7]. In a wide range of industrial fields, CAN bus can be taken as the field device-level communication bus and compared with other buses, which has high reliability and cost performance. This will be a main direction for CAN technology open for the development and applications of IoT [8]. CAN is a multimaster serial communication bus and according to its basic design specifications, it has a fast bit rate and high anti-electromagnetic interference, and it can detect any error. Firstly, the CAN controller works in a variety of modes, and each node in the network can adopt bit-by-bit arbitration of lossless structure and send data to the bus contention according to the bus access priority and message identifier. Besides, CAN protocol has replaced station address coding with communication data coding, enabling different nodes to receive the same data at the same time [9]. These characteristics have led

to strong real-time data communication among network nodes formed by CAN bus and the easiness to constitute a redundant structure and improved the reliability and flexibility of the system. The data communication of CAN bus has outstanding reliability, timeliness, and flexibility. Due to its excellent performance and unique design, CAN bus has attracted more and more attention. Most automobile manufacturers have adopted CAN bus to realize the data communication between internal control system and various test and execution institutions. Meanwhile, because of the characteristics of CAN bus, its application is no longer restricted to auto industry, but it also develops towards IoT and other fields. CAN has become an international standard and it has already been considered as one of the most promising field buses [10, 11].

In as early as 1998, Massachusetts Institute of Technology in the United States had come up with the idea of IoT at that time, i.e., EPC system, and one year later, the concept of IoT had been deemed as the logistics network based on Radio Frequency Identification (RFID) [12]. IoT has not had a formal definition: IoT is the network that makes all objects which have and can implement independent functions achieve interconnection, until 2005 when the International Telecommunication Union (ITU) released "ITU Internet Report 2005: Internet of Things" [13]. CAN bus is one of the field buses which have been applied the most in the world. It was first launched by Bosch, a German company, and used in the data connection protocols between internal measurement and execution units. The application of CAN bus ranges from high-speed networks to low-cost multiline network and it is widely applied in the data communication between various detection and execution agencies in the control system [14]. CAN bus can mount many communication nodes, the signals between which are transmitted via the bus so as to realize between-node communication. The use of RS-485 can only constitute a master-slave structural system and the communication can only be conducted by means of master station polling. Worse still, the system has poor timeliness and reliability [15].

In theory, as long as the bus has enough load, the number of nodes is not limited, and the repeater can improve the load. CAN bus can effectively reduce the number of wires and it has flexible communication, excellent timeliness, and high reliability. However, the communication quality of CAN is related to many factors, mainly including the distance between communication nodes, the type and impedance of communication lines, crystal oscillator error of various communication nodes, and the deviation caused by external environment [16]. In practical engineering field, the preset CAN sample point may not be able to guarantee the data receiving and dispatching quality of communication nodes and there may be frame loss and receiving and dispatching abnormalities in data which affect the normal operation of product functions due to the impact of the above factors. To solve the problems in existing technology, this paper provides an IoT-oriented scheme used in CAN bus message management and the hardware design method

of CAN bus interface circuit of highly integrated communication controller so as to guarantee the safety of network communication [17].

### 3. IoT-Oriented CAN Communication Control Design

**3.1. Key Issues in Interface Circuit Design.** Any node in CAN can take an initiative and send a message to other nodes in the network irrespective of master and slave at any moment. It adopts nondestructive bus arbitration technology. When two nodes send a message to the network at the same time, the node with low priority voluntarily stops data transmission while the one with high priority continues to transmit the data insusceptibly. In addition, it has the functions of point-to-point, one point to multipoints, and global broadcast data transmission and reception. Every frame of message has CRC check and other error detection measures with an extremely low error rate and high reliability. When a serious mistake occurs in message transmission, the node can automatically cut its connection with the bus so that other operations in the bus are not affected [18].

**3.1.1. Optoelectronic Isolation Circuit.** Although optoelectronic isolation circuit can strengthen the antijamming capability of the system, it also increases the propagation delay time of CAN bus effective loop signal and decrease the communication speed or distance. CAN transceiver itself has instantaneous anti-interference, radio frequency interference (RFI), and the ability to achieve thermal protection; additionally, its current limit circuit has also provided further protection to the bus. Therefore, in case of short field transmission distance and little electromagnetic interference, optoelectronic isolation may not be adopted to make the system reach the maximum communication speed or distance and simplify the interface circuit. If optoelectronic isolation is needed in the field environment, high-speed optoelectronic isolation device should be selected to reduce the propagation delay time of effective loop signal of CAN bus. The propagation delay time is short and it is approximate to the level of TTL circuit propagation delay time [5].

**3.1.2. Power Isolation.** The power  $V_{dd}$  and  $V_{cc}$  used in the two sides of optoelectronic isolation device must be completely isolated. Otherwise, optoelectronic isolation will not play its due role. Power isolation can be achieved through low-power DC/DC power isolation module.

**3.1.3. Bus Impedance Matching.** The terminal of CAN bus must be connected with  $n$  resistors, which have a significant role in bus impedance matching; otherwise, it will greatly reduce the reliability and anti-interference performance of bus data communication and it may even make it impossible to carry out communication. In order to improve the antijamming capability of interface circuit, the following measures can also be taken into consideration: connect  $n$

parallel capacitors between CANH and CANL of the transceiver and the earth to filter the high-frequency interference in the bus and prevent electromagnetic radiation [19].

**3.2. CAN Bus Synchronization Mechanism.** CAN communication is an asynchronous serial communication. Its synchronization method is not achieved by providing extra clock signals but by realizing read-write synchronization through bit timing, fixed message frame structure, and hard synchronization and resynchronization operations. The condition for communication between various nodes in CAN bus network is that they have the same normal bit rate (NBR), i.e., the bits are transmitted every second. Its reciprocal is called normal bit time  $t_{bit}$  which is composed of four nonoverlapping time periods: reciprocal period, transmission period, phase buffer period 1, and phase buffer period 2. The normal bit time can be defined as follows:

$$t_{bit} = \frac{1}{NBR} = t_{SyncSeg} + t_{propSeg} + t_{PS1} + t_{PS2}. \quad (1)$$

Every period of time in normal bit time is made up of multiple time quanta TQ, the length of which depends on the frequency of oscillator  $f_{OSC}$ . The relation between them is as follows:

$$TQ = \frac{2BRP}{f_{OSC}}. \quad (2)$$

In which, BRP is the Baud rate prescale coefficient and for the device that adopts the same clock frequency of the master oscillator in CAN bus network, the normal bit rate NBR can be adjusted to the same numerical value only by setting BRP and the TQ in every time period [20, 21].

When IoT conducts wireless communication and if the transport layer has a high packet loss speed-up rate or the data is blocked, the channel protection mechanism starts automatically. Assume that  $S(v_j, v_{DAPi})$  is the link communication of transmit data and  $v_0, v_{DAPi}$  represents the data source and destination aggregation node. In other words, when the system accesses any aggregation point, stop the transmission if at the time of  $t_0$  and the forward node  $v_i$  in the congestion point will find the homologous data continuously accumulated in the buffer zone and meanwhile conduct self-inspection and abandon. The backward node will not delete the link in the transport layer. So, there is no homologous data in the buffer zone. Therefore, the node can make a judgment according to whether the connection channel is congested or not so as to start the multiagent dynamic additional channel protection mechanism [22, 23].

With regards to protection mechanism, the node goes along the transmission channel in this technology. During the traverse, the proxy includes two parameters in composite measurement: the vacancy and surplus bandwidth rate of the receiving node. In the composite measurement, the corresponding equation to calculate the queue occupancy rate of the receiving node is as follows:

$$\mu_k(t) = 1 - \frac{q_k - \int_{t_{0-}}^{t_0} v_k^1(t) dt + \int_{t_0}^{t_{0+}} v_k^x(t) dt}{B_k} + \xi. \quad (3)$$

In which,  $\mu_k$  is the receiving node,  $v_k$  is the vacancy of queue,  $q_k$  is the current queue length of node  $v_k$ , and  $v_k^1(t)$  and  $v_k^x(t)$  represent the velocities for communication buffer zone of the node at  $t$  phase to send and receive grouping, respectively.  $\xi$  is the buffer zone of the queue and it is used to control the transmission of “flow” over message and  $\xi \in [0.01, 0.08]$ .  $B_k$  represents the size of the communication buffer zone of node  $v_k$  [24].

The composite measurement of proxy travel restoration is defined as follows:

$$M(e_{hk}) = a\mu(t) + b\gamma(e_{hk}) + 0.5. \quad (4)$$

In which,  $\alpha$  and  $\beta$  represent weight coefficient and  $a + b = 1$ . It controls the transmission of real-time message through low  $a/b$  rate as data flow will generate low jitter during the transmission. So, the rate shall increase properly.

There are two main types of CAN physical layer. The CAN communication network in Figure 1 is a high-speed, short-distance “closed-loop network” that complies with the ISO11898 standard. Its maximum bus length is 40 m, and its communication speed is up to 1 Mbps. Each end requires a 120  $\Omega$  resistor [25].

Figure 2 is a low-speed, long-distance “open-loop network” that follows the ISO11519-2 standard. Its maximum transmission distance is 1 km, and its maximum communication rate is 125 kbps. The two buses are independent and do not form a closed loop. Each bus is required. There is a 2.2 k $\Omega$  resistor in series on each [26].

## 4. Experimental Procedure

**4.1. Create the Message Transmission Model.** This section is about creating the part which transmits CAN message in the model. It sets and connects with each module by using the relevant modules in Simulink library.

**4.1.1. Create the Model.** Start Simulink in MATLAB software platform, select “Blank Model”  $\rightarrow$  “Create Model,” and complete the creation of a new blank model window.

**4.1.2. Add Related Modules.** Click the button of “Library Browser” in the toolbar and open Simulink module library. In the left side is the available module library. Select the related modules from “Vehicle Network Toolbox”  $\rightarrow$  “CAN Communication” and use them in the transmission of CAN communication in this instance.

Move the modules of “CAN Configuration,” “CAN Pack,” and “CAN Transmit” to the editing interface, respectively, and create one instance for each of them. The related module descriptions are shown in Table 1.

Select different types of modules as data source. In this paper, constant is used as the data source to verify the CAN communication model.

**4.1.3. Connect Modules.** The above modules are connected through wires: the output port of Constant is connected with the input port of CAN Pack, the output port of which is connected to the input port of Transmit and CAN Configuration is not connected with any other modules. In this module configuration, CAN channel is used for communication.

**4.1.4. Set Module Parameters.** Double click CAN configuration to configure its parameters. The specific parameter description is shown in Table 2.

Double click CAN Pack to configure its parameters, the specific descriptions of which are shown in Table 3.

Double click CAN Transmit to configure its parameters, the specific description of which can be found in Table 4.

The model structure after various modules are connected and configured in the transmission part is shown in Figure 3.

**4.2. Create the Model to Receive Message.** This section is used to create the part which receives CAN message in the model. It sets and connects various modules by using related modules in Simulink library.

**4.2.1. Add Related Modules.** Select related modules from “Vehicle Network Toolbox”  $\rightarrow$  “CAN Communication” and use them to receive CAN communications in this instance. Select the module of “Simulink”  $\rightarrow$  “Sinks”  $\rightarrow$  “Scope” to display the data received.

Move the modules: “CAN Configuration,” “CAN Receive,” “CAN Unpack,” and “Scope” to the editing interface, respectively, and create one instance for each. The related module descriptions are shown in Table 5.

**4.2.2. Connect Modules.** Create CAN message pack subsystem with the module of “Function-Call Subsystem” in “Ports & Subsystems” module library and move CAN Unpack to this subsystem.

In CAN Receive, CAN Msg (output port) is connected with In1 (input port) of CAN Unpack subsystem [27]. In CAN Receive,  $f()$  (output port) is connected with function() (input port) of CAN Unpack subsystem and in CAN Unpack subsystem, CAN Msg(output port) is connected with the input port of scope [28]. CAN Configuration1 is not connected with any other modules and it configures CAN channel for communication [29, 30].

**4.2.3. Set Module Parameters.** Double click CAN Configuration1 to configure its parameters and their descriptions are indicated in Table 6.

Double click CAN Receive to configure its parameters. The related descriptions of these parameters are demonstrated in Table 7.

Double click CAN Unpack to configure its parameters. Refer to CAN Pack for the related parameters’ description.

The model structure of the reception part after various modules are connected and configured is shown in Figure 4.

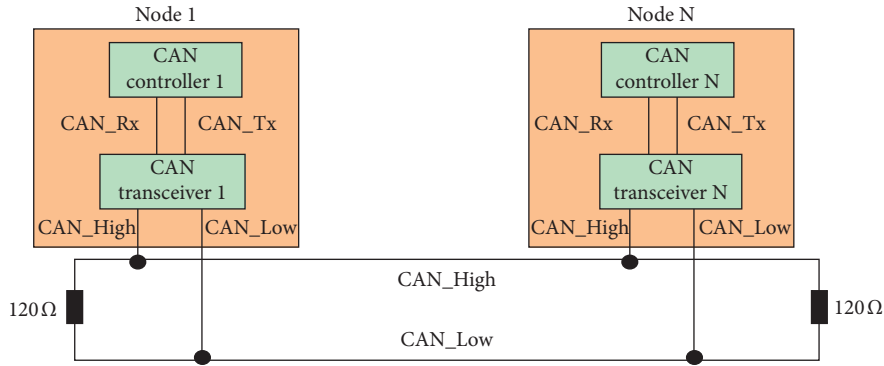


FIGURE 1: CAN closed-loop bus communication network.

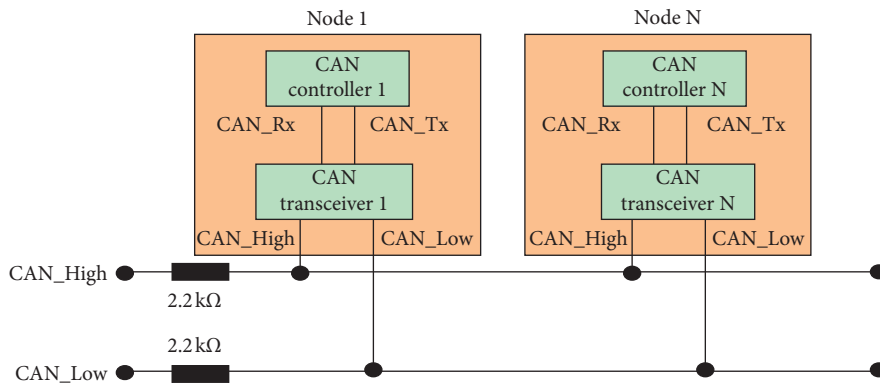


FIGURE 2: CAN open-loop bus communication network.

TABLE 1: Description of modules in CAN communication transmission.

| Module            | Diagram | Description   |
|-------------------|---------|---|
| CAN Configuration |         | It is used to configure the parameters of CAN device which is used in transmitting and receiving message and it uses one CAN Configuration module to configure every device that sends and receives message in the model. |
| CAN Pack          |         | It is used to upload signal data at a designated interval to a message during the simulation. The number of module inputs depends on the dynamic adjustment of the number of specified signals. It has an output port.    |
| CAN Transmit      |         | It is used to send and transmit data to the virtual CAN channel, and it can transmit a single message or message array. There is no output port in this module.   |

TABLE 2: Description of parameters of CAN Configuration.

| Parameter        | Value                           | Description   |
|------------------|---------------------------------|---|
| Device           | MathWorks Virtual 1 (Channel 1) | Select from list CAN device and its necessary channel. The device is used to transmit messages.   |
| Bus speed        | 500,000                         | Set the bus speed attribute for the device selected above with bits per second as the unit.   |
| Acknowledge mode | Normal                          | Check whether the designated channel is in normal mode (normal) or silent mode (silent). If normal, the channel can normally both receive and transmit message while if silent, the channel can only receive message. |

TABLE 3: Description of parameters of CAN pack.

| Parameter        | Value                        | Description   |
|------------------|------------------------------|---|
| Data is input as | Raw data                     | It selects the input signal. The raw data input the data in the form of uint8 vector array; manually specified signals can assign the definition of data signal; and CANdb specified signals can assign the CAN database file which includes the definitions of message and signal. |
| Name             | CAN Msg                      | The name of CAN message can be assigned with the default value of CAN Msg.  |
| Identifier type  | Standard (11-bit identifier) | It assigns whether the identifier of CAN message is standard or extended.   |
| Identifier       | 500                          | It assigns the ID of CAN message. For standard identifier, the figure must be any positive integer from 0 to 2047; and for extended identifier, it must be any positive integer from 0 to 536870911.  |
| Length (bytes)   | 8                            | It assigns the length of CAN message, the scope of which ranges from 0 to 8 bytes.  |

TABLE 4: Description of parameters of CAN transmit.

| Parameter | Value                           | Description  |
|-----------|---------------------------------|--|
| Device    | MathWorks Virtual 1 (Channel 1) | It selects the CAN device and channel which are used to transmit CAN message to the network. |

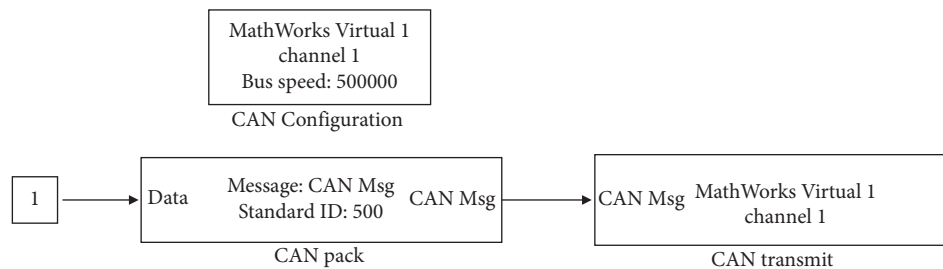


FIGURE 3: Transmission part model structure.

TABLE 5: Description of modules in the reception part of CAN communication.

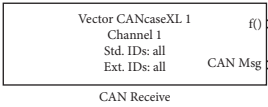
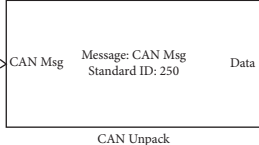

| Module      | Diagram   | Description   |
|-------------|---|---|
| CAN Receive |  | It is used to receive messages from CAN and transmit them to Simulink model. In this module, there are two output ports: $f()$ is a trigger to function-call subsystem and CAN Msg includes the CAN messages received during the specific timestep. |
| CAN Unpack  |  | It is used to unpack CAN message by using designated output parameters in every timestep into signal data and the data is output as a single signal.  |
| Scope       |  | It is used to display time-domain signal.   |

TABLE 6: Description of parameters of CAN Configuration1.

| Parameter | Value                           | Description  |
|-----------|---------------------------------|--|
| Device    | MathWorks Virtual 1 (Channel 2) | Select from list CAN device and the necessary channel and use this device to receive messages. |

4.3. *Operate the Model.* Select in the Toolbar the button of “Run” and click it to simulate. When operating simulation, CAN Transmit will obtain message from CAN pack. Then it will transmit the message through virtual channel 1 and the

CAN Receive in virtual channel 2 will receive this message and transmit it to CAN Unpack subsystem to unpack the message. Finally, double click scope to check the transmission result.



TABLE 7: Description of parameters of CAN Receive.

| Parameter                                    | Value                              | Description   |
|--|------------------------------------|---|
| Device                                       | MathWorks Virtual 1<br>(Channel 2) | It is used to select CAN device and channel to receive CAN message from this device.  |
| Sample time                                  | 0.01                               | It is used to assign the sample time of modules during the simulation and it is defined as the operating frequency of CAN receive during this time.   |
| Number of messages received at each timestep | All                                | It is used to select the number of messages the module receives at each designated timestep. All is to transmit all messages in the buffer zone to model and 1 is to transmit one message to the model. |

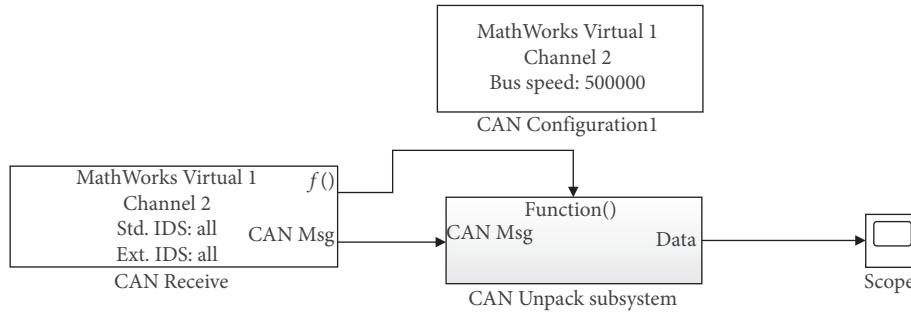


FIGURE 4: Receiving part model structure.

## 5. Simulation Experiment and Results Analysis

After the above links, the model structure which includes such links as transmission, reception, and operation are basically built. In this section, different types of modules will be selected as data sources so as to verify the CAN communication Simulink model.

Use “Constant” as the data source and add a Constant to the model from Simulink → Sources. Constant generates real-number or plural constant-value signal, and it is used to provide the input of constant signal. Among them, the parameter of “Constant value” is set as [1 2 3 4 5 6 7 8]. The final output result of CAN communication is shown in Figure 5.

Use “Sine Wave” as the data source and add a Sine Wave to the model from Simulink → Sources library. Sine Wave takes simulation time as the time source to generate sine wave and outputs sinusoid wave. The equation of wave function is

$$y = \text{amplitude} \times \sin(\text{frequency} \times \text{time} + \text{phase}) + \text{bias}. \quad (5)$$

In this instance, the value of each parameter is set as follows: amplitude is 1, bias is 200, frequency is 1, and phase is 0. The final output result of CAN communication is shown in Figure 6.

Use “Repeating Sequence” as the data source and add a Repeating Sequence to the model from Simulink → Sources library. Repeating Sequence is used to generate periodic signal of any shape and it can output periodic scalar signal with wave shape designated by parameters: time values and output values. Among them, Time values assign the output time vector and Output values specify the corresponding signal amplitude vector to output time. In this instance, the

values of various parameters are set as follows: Time values within [0 5], and Output values within [0 100]; in other words, starting from simulation, repeat one sawtooth waveform with the maximum amplitude of 100 every 5 seconds. The final output result of CAN communication is indicated in Figure 7.

Use “Step” as the data source and add a Step Module to the model from Simulink → Sources library. Step is used to generate step function and it can provide 2 steps which can define level during the designated time. If the simulation time is less than the parameter value of Step time, the parameter value of Initial value will be the output and if it is bigger than or equal to Step time, the output is the parameter value of Final value. In this instance, various parameters are set as follows: step time is 15, initial value is 5, and final value is 100. The final output result of CAN communication is shown in Figure 8.

Use “Pulse Generator” as the data source and add a module of Pulse Generator to the model from Pulse Generator library. Pulse Generator is used to generate square-wave pulse at a fixed interval and its waveform parameters, Amplitude, Pulse Width, Period, and Phase delay, and determine the shape of output waveform. In this instance, the values of various parameters are set as follows: phase delay is 0, amplitude is 100, pulse width is 20, and period is 5. The final output result of CAN communication is shown in Figure 9.

Take “Ramp” as the data source and add a “Ramp” module to the model from Simulink → Sources library. Ramp is used to generate the signals that constantly increase or decrease, and it can generate the signal which changes at a designated speed starting from designated time and value. Its parameters of slope, start time, and initial output decide the features of output signal. In this instance, the values of these

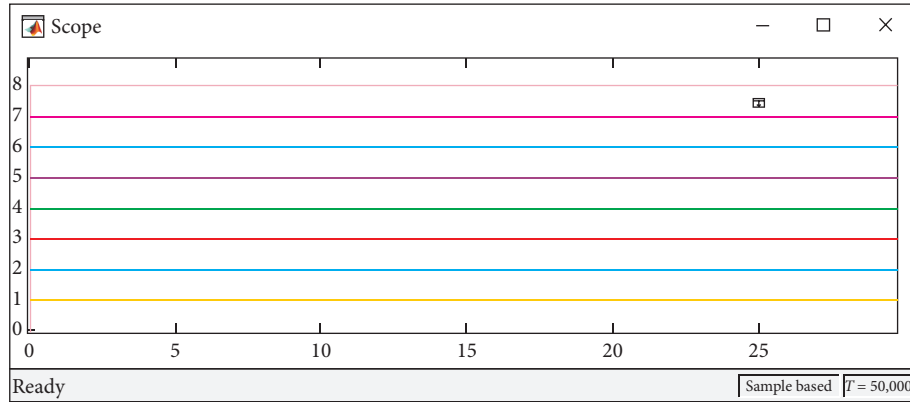


FIGURE 5: Output result of CAN communication in constant.

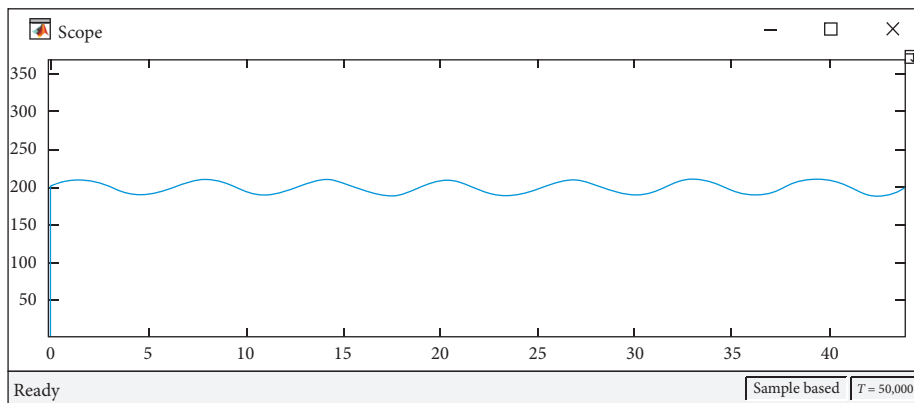


FIGURE 6: Output result of CAN communication of sine wave.

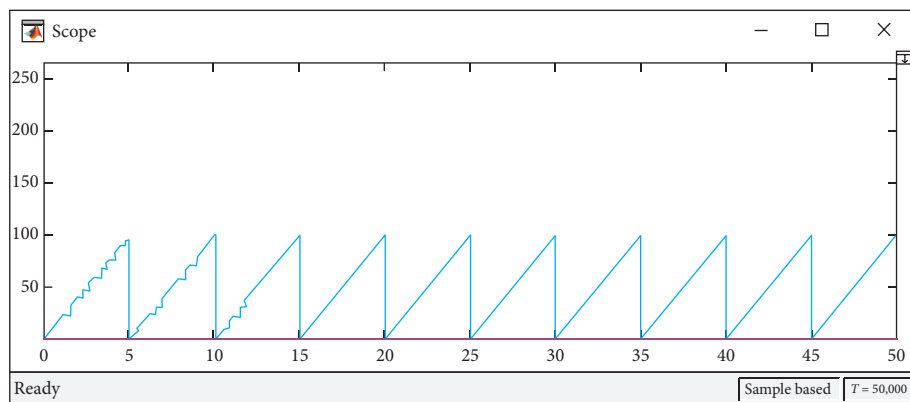


FIGURE 7: Output result of CAN communication of repeating sequence.

parameters are set as follows: slope is 5, start time is 15, and initial output is 5. The final output result of CAN communication is shown in Figure 10.

Use “Signal Generator” as the data source and from Simulink → Sources library, add a model of Signal Generator to the model. Signal Generator can be used to generate four different waveforms: sine wave, square wave, sawtooth wave, and random wave. The equation of wave function is

$$Y(t) = \text{Amp} * \text{Waveform}(\text{Freq}, t). \quad (6)$$

In this instance, sine wave is selected, and the values of parameters are set as follows: amplitude is 1 and frequency is 1. The final output result of CAN communication is shown in Figure 11.

Use “Chirp Signal” as the data source and from Simulink → Sources library, add a “Chirp Signal” to the model.

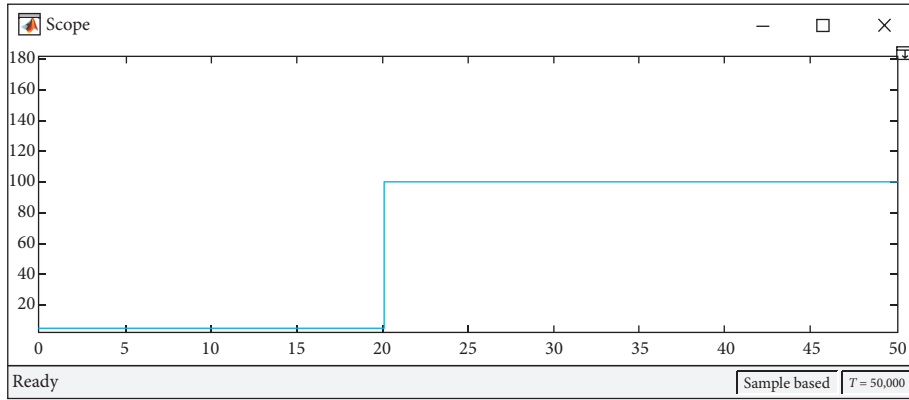


FIGURE 8: Output result of CAN communication of step.

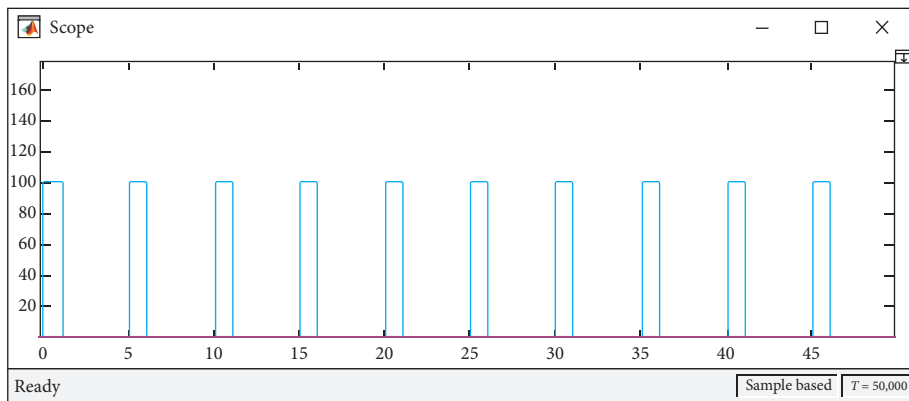


FIGURE 9: Output result of CAN communication of pulse generator.

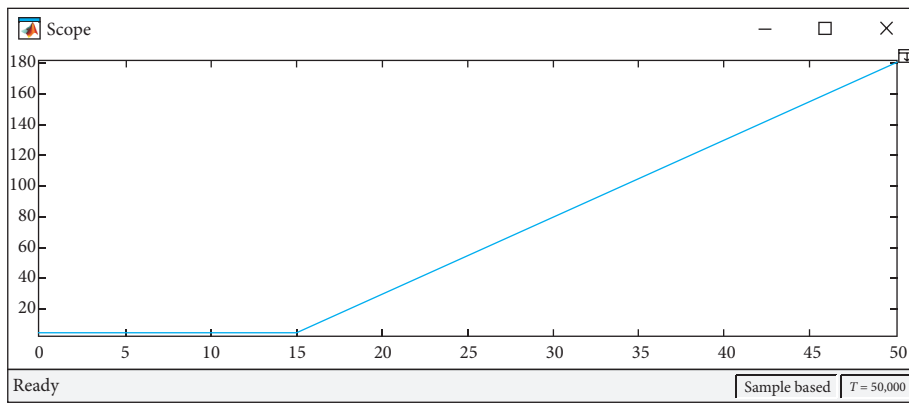


FIGURE 10: Output result of CAN communication in ramp.

This module is used to generate the waves, the frequency of which constantly increases. The parameters of initial frequency, target time, and frequency at target time decide its output. In this instance, the values of parameters are set as follows: initial frequency is 0.1, target time is 100, and

frequency at target time is 5. The final output result of CAN communication is shown in Figure 12.

Through the above experimental results, it can be observed that the IoT-oriented network communication model not only has low consumption, but also solves the front-end

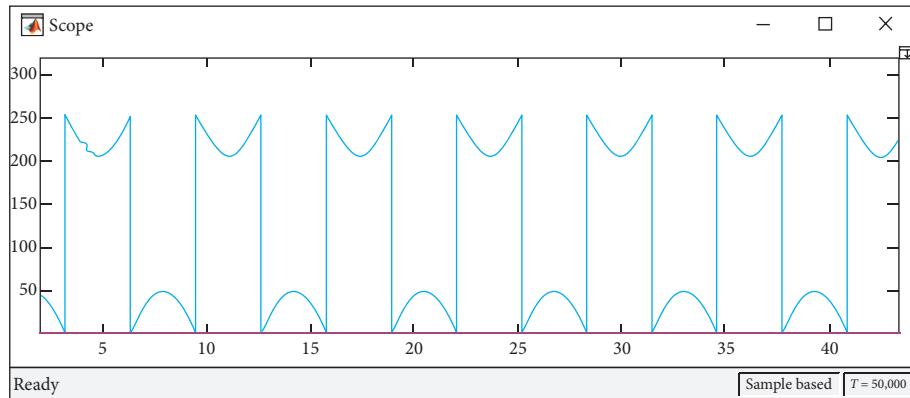


FIGURE 11: Output result of CAN communication of signal generator.

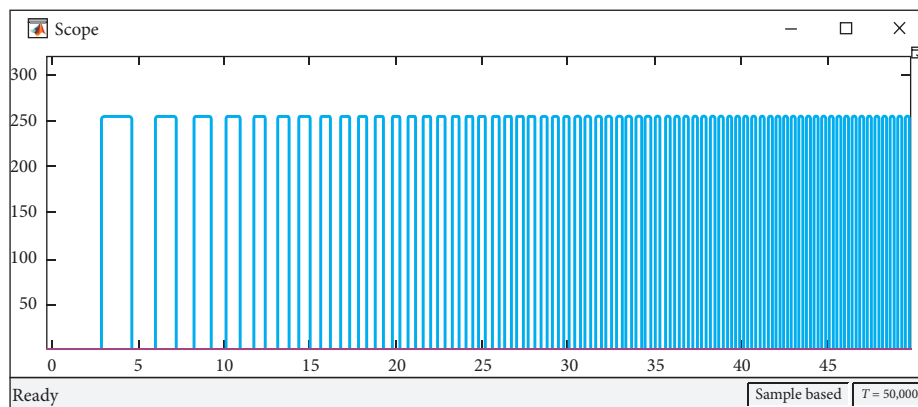


FIGURE 12: Output result of CAN communication of chirp signal.

communication compatibility issues caused by different communication protocols.

## 6. Conclusions

IoT is an extension of the Internet, which extends to the communication network between objects through wireless network. It goes through various sensor devices and it can perform real-time collection, remote monitoring, connection operations, and interaction information on the device. It can constitute a new-type network with Internet technology. Interface circuit is an important link in CAN bus network and its reliability and security have directly affected the operation of the entire communication network. This paper has summarized several key issues to be noted in the design of CAN interface circuit, grasped the key in the design, and proposed an IoT-oriented universal CAN bus design, which has enhanced the quality and performance of multi-interface circuit and ensured the secure and reliable operation of CAN bus. The simulation experiment has proven that the proposed design scheme is effective.

## Data Availability

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also form part of an ongoing study.

## Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was supported by the Jiangsu Basic Research Program (Natural Science Foundation) Project (Grant no. BK20191393), and Jiangsu Agriculture Science and Technology Innovation Fund (JASTIF, CX(20)3038).

## References

- [1] C.-Z.-E. Li and Z. W. Deng, "The embedded modules solution of household Internet of Things System and the future development," *Procedia Computer Science*, vol. 166, pp. 350–356, 2020.
- [2] X. Huang, "Intelligent remote monitoring and manufacturing system of production line based on industrial Internet of Things," *Computer Communications*, vol. 150, no. 15, pp. 421–428, 2020.
- [3] B. Diène, J. J. P. C. Rodrigues, O. Diallo, E. H. M. Ndoeye, and V. V. Korotaev, "Data management techniques for internet of things," *Mechanical Systems and Signal Processing*, vol. 138, no. 8, p. 106564, 2020.
- [4] J. H. Nord, A. Koohang, and J. Paliszkiwicz, "The internet of things: review and theoretical framework," *Expert Systems with Applications*, vol. 133, no. 1, pp. 97–108, 2019.

- [5] M. K. Ishak and F. K. Khan, "Unique message authentication security approach based controller area network (CAN) for anti-lock braking system (ABS) in vehicle network," *Procedia Computer Science*, vol. 160, pp. 93–100, 2019.
- [6] X. Kong, Y. Zheng, M. Ouyang et al., "Signal synchronization for massive data storage in modular battery management system with controller area network," *Applied Energy*, vol. 197, no. 1, pp. 52–62, 2017.
- [7] Y. Jiang, "Combination of wearable sensors and internet of things and its application in sports rehabilitation," *Computer Communications*, vol. 150, no. 15, pp. 167–176, 2020.
- [8] S. Mubeen, J. Mäki-Turja, and M. Sjödin, "Integrating mixed transmission and practical limitations with the worst-case response-time analysis for controller area network," *Journal of Systems and Software*, vol. 99, no. 1, pp. 66–84, 2015.
- [9] M. B. N. Shah, A. R. Husain, S. Punekkat, and R. S. Dobrin, "A new error handling algorithm for controller area network in networked control system," *Computers in Industry*, vol. 64, no. 8, pp. 984–997, 2013.
- [10] Y. Shoukry, H. Shokry, and S. Hammad, "Distributed dynamic scheduling of controller area network messages for networked embedded control systems," *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 1959–1964, 2011.
- [11] S. Mubeen, J. Mäki-Turja, and M. Sjödin, "MPS-CAN analyzer: integrated implementation of response-time analyses for controller area network," *Journal of Systems Architecture*, vol. 60, no. 10, pp. 828–841, 2014.
- [12] M. S. Manshahia, "Grey wolf algorithm based energy-efficient data transmission in internet of things," *Procedia Computer Science*, vol. 160, pp. 604–609, 2019.
- [13] S. Zeadally, F. K. Shaikh, A. Talpur, and Q. Z. Sheng, "Design architectures for energy harvesting in the Internet of Things," *Renewable and Sustainable Energy Reviews*, vol. 128, no. 8, p. 109901, 2020.
- [14] G. Cena, I. Cibrario Bertolotti, T. Hu, and A. Valenzano, "On a family of run length limited, block decodable codes to prevent payload-induced jitter in Controller Area Networks," *Computer Standards & Interfaces*, vol. 35, no. 5, pp. 536–548, 2013.
- [15] M. Park and X. Piao, "Enhanced utilization bound of rate-monotonic scheduling in controller area networks," *Information Processing Letters*, vol. 114, no. 11, pp. 603–606, 2014.
- [16] S. M. Nekooei, G. Chen, and R. K. Rayudu, "Automatic design of fuzzy logic controllers for medium access control in wireless body area networks—an evolutionary approach," *Applied Soft Computing*, vol. 56, no. 7, pp. 245–261, 2017.
- [17] X. Zhu, H. Zhang, D. Cao, and Z. Fang, "Robust control of integrated motor-transmission powertrain system over controller area network for automotive applications," *Mechanical Systems and Signal Processing*, vol. 58–59, pp. 15–28, 2015.
- [18] S. E. Marx, J. D. Luck, S. K. Pitla, and R. M. Hoy, "Comparing various hardware/software solutions and conversion methods for Controller Area Network (CAN) bus data collection," *Computers and Electronics in Agriculture*, vol. 128, no. 10, pp. 141–148, 2016.
- [19] W. Wei, H. Song, W. Li, P. Shen, and A. Vasilakos, "Gradient-driven parking navigation using a continuous information potential field based on wireless sensor network," *Information Sciences*, vol. 408, no. 2, pp. 100–114, 2017.
- [20] Y. Tao, L. Dong, and L. Chai, "Design of special communication protocol for electric rudder system based on CAN bus," *Aviation Precision Manufacturing Technology*, vol. 53, no. 6, pp. 22–26, 2017.
- [21] G. Breaban, M. Koedam, S. Stuijk, and K. Goossens, "Time synchronization for an emulated CAN device on a multi-processor system on chip," *Microprocessors and Microsystems*, vol. 52, pp. 523–533, 2017.
- [22] C.-Y. Wan, A. T. Campbell, and L. Krishnamurthy, "Pump-slowly, fetch-quickly (PSFQ): a reliable transport protocol for sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 862–872, 2005.
- [23] Y. Sun, Y. Weng, B. Luo et al., "Gesture recognition algorithm based on multi-scale feature fusion in RGB-D images," *IET Image Processing*, 2020.
- [24] Q. Pu and X. Wu, "Research on wireless routing algorithm based on congestion control," *Communication Technology*, vol. 49, no. 3, pp. 312–317, 2016.
- [25] R. Alaei, P. Moallem, and A. Bohlooli, "Statistical based algorithm for reducing bit stuffing in the controller area networks," *Microelectronics Journal*, vol. 101, no. 7, p. 104794, 2020.
- [26] J. Bauer, R. Helmke, A. Bothe, and N. Aschenbruck, "CAN't track us: adaptable privacy for ISOBUS controller area networks," *Computer Standards & Interfaces*, vol. 66, no. 10, p. 103344, 2019.
- [27] Z. Chen, D. Chen, Y. Zhang, X. Cheng, M. Zhang, and C. Wu, "Deep learning for autonomous ship-oriented small ship detection," *Safety Science*, vol. 130, p. 104812, 2020.
- [28] L. Dong, W. Wu, Q. Guo, M. N. Satpute, T. Znati, and D. Z. Du, "Reliability-aware offloading and allocation in multilevel edge computing system," *IEEE Transactions on Reliability*, pp. 1–12, 2019.
- [29] R. Ma, L. Zhang, G. Li, D. Jiang, S. Xu, and D. Chen, "Grasping force prediction based on sEMG signals," *Alexandria Engineering Journal*, vol. 59, no. 3, p. 1135, 2020.
- [30] W. Wei, X. Fan, H. Song, and J. Yang, "Imperfect information dynamic stackelberg game based resource allocation using hidden Markov for cloud computing," *IEEE Transactions on Services Computing*, vol. 11, no. 1, pp. 78–89, 2016.

## Research Article

# Heuristic Sensing: An Uncertainty Exploration Method in Imperfect Information Games

Zhenyang Guo <sup>1</sup>, Xuan Wang <sup>1</sup>, Shuhan Qi <sup>1,2</sup>, Tao Qian <sup>1</sup> and Jiajia Zhang <sup>1,2</sup>

<sup>1</sup>Harbin University of Technology Shenzhen, Shenzhen 518055, China

<sup>2</sup>Pingan-Hitsz Intelligence Finance Research Center, Shenzhen 518055, China

Correspondence should be addressed to Jiajia Zhang; zhangjiajia@hit.edu.cn

Received 26 August 2020; Revised 6 October 2020; Accepted 9 October 2020; Published 24 October 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Zhenyang Guo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Imperfect information games have served as benchmarks and milestones in fields of artificial intelligence (AI) and game theory for decades. Sensing and exploiting information to effectively describe the game environment is of critical importance for game solving, besides computing or approximating an optimal strategy. Reconnaissance blind chess (RBC), a new variant of chess, is a quintessential game of imperfect information where the player's actions are definitely unobserved by the opponent. This characteristic of RBC exponentially expands the scale of the information set and extremely invokes uncertainty of the game environment. In this paper, we introduce a novel sense method, Heuristic Search of Uncertainty Control (HSUC), to significantly reduce the uncertainty of real-time information set. The key idea of HSUC is to consider the whole uncertainty of the environment rather than predicting the opponents' strategy. Furthermore, we realize a practical framework for RBC game that incorporates our HSUC method with Monte Carlo Tree Search (MCTS). In the experiments, HSUC has shown better effectiveness and robustness than comparison opponents in information sensing. It is worth mentioning that our RBC game agent has won the first place in terms of uncertainty management in NeurIPS 2019 RBC tournament.

## 1. Introduction

Game theory is the mathematical study of interaction among independent, self-interested players, providing a very simple but powerful paradigm to capture decision problem. The classical category divides games into perfect information games (PIGs) and imperfect information games (IIGs). In PIGs, players can obtain complete information of game environment. However, pervasively existing in real world, players cannot sense complete or reliable information of games. IIGs address these cases and model strategic interactions among agents with only partial or unreliable information. Thus, the exploitation of imperfect information of IIGs is one of the most critical challenges for game solving since how well players understand the game environment greatly influences the effectiveness of their strategies.

In this paper, we focus on a recently introduced IIG, reconnaissance blind chess for research of imperfect information exploitation. Reconnaissance blind chess is actually a

family of games, and we only focus on one variant, which we will refer to as RBC for simplicity [1]. RBC was designed intentionally to add a certain amount of uncertainty by adjusting some rules of chess and adding an explicit sense step.

Furthermore, we utilize the algorithm of MCTS in this paper. Monte Carlo (MC) method has been used extensively in PIGs [2] and IIGs [3] which uses random simulations to approximate the true value of states in IIGs. Furthermore, Upper Confidence Bound for Trees (UCT) is the most popular MCTS algorithm, using upper confidence bounds, a formula trying to settle the exploitation-exploration dilemma, as a tree policy for selection and expansion [4]. UCT converges to Minimax, the optimal algorithm used for two-player zero-sum games [5], given enough time and memory. However, the reality is that time and memory are limited. Hence, one severe challenge of the MCTS + UCT structure is the contradiction between the accuracy of states' estimation and limited simulation time both of which are critical for a competitive game program.

The contribution of this paper is twofold. First, we introduce a novel sense method, HSUC, to effectively exploit and manage the uncertainty of the game environment. Our method is no longer entirely dependent on the accuracy of opponents’ actions prediction which severely relies on plenty of simulation time. Instead of that, the key idea of HSUC is focusing on reducing the whole uncertainty of the environment, which is characterized by real-time information set in the game solving process. Second, we realize a practical framework for RBC game that incorporates HSUC with MCTS + UCT. NeurIPS 2019 tournament contains a final win rate rank and several ranks of different indicators. Our agent constructed with this framework has ranked the 7<sup>th</sup> in the final rank and won the first rank in terms of uncertainty management in particular.

## 2. Environment and Preliminaries

In this part, we will briefly introduce the rules of RBC, explain why RBC is a problem worth studying, and point out the difficulty of research. And then, we provide some preliminaries for later discussion.

*2.1. Environment: RBC and Its Challenges.* The major difference between RBC and standard chess is that RBC players are not informed of the opponent’s actions in the process of the game. For managing this hidden information, an additional step called “sense” is embedded prior to the move step. During the sense step, a player selects a square of the chessboard and learns all pieces and their types within the square, and that action is invisible to another player. This step is the most important way for the player to obtain real information about the opponent. That means players should consider their sense strategies to choose a region to review an unknown part of the board. In addition, some changes have been made to other rules, for example, the player wins by capturing the opponent’s king, but, in chess, a win occurs when the King is in under attack or in “check” and every possible move by the King will also put it in check. Since the player cannot see the opponent’s chess pieces, some invalid actions may occur when the player moves. For details, please check the description of the website (<https://rbc.jhuapl.edu/gameRules>). The game tree of RBC is shown in Figure 1.

The past decades have witnessed rapid progress in the ability of AI systems to play increasingly complex games, such as go of PIGs [6] and poker of IIGs [7]. Not long ago, Brown et al. proposed poker agent Pluribus to solve the problem of multiplayer poker [8]. But RBC, as an IIG, is even more complex in certain aspects than multiplayer poker. We will discuss challenges in RBC in the following two aspects: the game size and the number of possible states in the information set.

Generally speaking, the game size can be measured by the number of states that players may encounter in the game. A practical method to measure the game size proposed by Shannon in 1950 [9] is widely adopted. According to the method, the game size of Lim 2-P Poker (Lim 2-P Poker

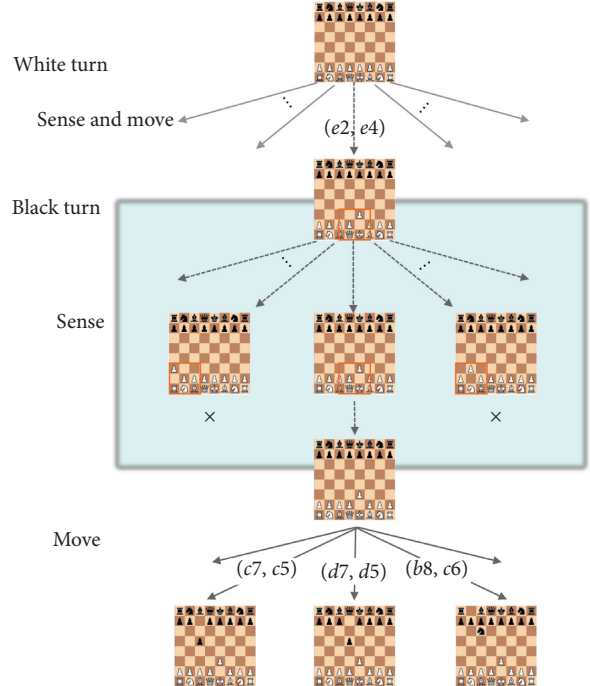


FIGURE 1: Game tree of RBC. One turn in RBC containing two phases cannot be observed by the opponent: sense and move. In the sense phase, a sense square (area in the red box) is chosen based on sense strategy firstly. Then, positions in the sense square of the chessboard are checked and conflicting states are removed from the information set based on sense result. After sense, a move is selected based on the move strategy. Move action (c7, c5) means that moving the piece of pawn from positions c7 to c5 and (b8, c6) means moving the knight in b8 to c6.

refers to Limit Heads-Up Texas Hold’em) is  $10^{13}$ , the game size of chess is  $10^{43}$ , and that of RBC is  $10^{139}$  [10]. Table 1 lists the number of states for several representative games and it denotes that RBC can approximately achieve a level similar to No-Limit Poker and go in terms of game size.

IIGs’ complexity can be measured by another metric: the average number of possible states in the information sets. In RBC, this metric represents how difficult it is to evaluate a given perceived state. Poor sense strategy may lead to an exponential growth of the scale of information sets. Thus, the key property of RBC is the information asymmetry, that is, the uncertainty about the opponent’s information. Table 2 shows that Jared Markowitz et al. [10] have calculated that the approximate average number of states of real-time information set in RBC is  $1.3 \times 10^{68}$ , which is even larger than Six-Player-No-Limit Poker.

### 2.2. Preliminaries

*2.2.1. Extensive-Form Games.* Sequential games are normally formalized as extensive-form games in which one or more agents or players perform sequential interactions. The extensive-form game can be described as a conceptual mode of six-tuple  $\langle P, H, Z, A_m(h), \sigma_c(h, a), u_p(z) \rangle$ :

- (1)  $P$ : the set of players.

TABLE 1: Approximate size of games.

| Game          | Size       |
|---------------|------------|
| Lim 2-P Poker | $10^{13}$  |
| Chess         | $10^{43}$  |
| RBC           | $10^{139}$ |
| No-Lim Poker  | $10^{162}$ |
| Go (19×19)    | $10^{170}$ |

TABLE 2: Approximate means the number of possible opponents' states in an information set.

| Game                | Game states          |
|---------------------|----------------------|
| Heads-Up No-Limit   | 1083                 |
| Chess               | 1                    |
| RBC                 | $1.3 \times 10^{68}$ |
| Six-Player-No-Limit | $6.4 \times 10^{14}$ |
| Go (19×19)          | 1                    |

- (2)  $H$ : a finite set  $H$  of sequences, the possible histories of actions, such that the empty sequence is in  $H$  and every prefix of a sequence in  $H$  is also in  $H$ .
- (3)  $Z$ : the set of all terminal states, corresponding to all leaf nodes in the game tree.
- (4)  $A(h)$ : the set of legal actions from state  $h$  ( $h \notin z$ ), corresponding to all edges starting from node  $h$  in the game tree.
- (5)  $\sigma_c(h, a)$ : the probability that chance will take action  $a \notin A(h)$  from state  $h$ .
- (6)  $u_p(z)$ : the payoff for player  $p$  if the game ends in state  $z$  ( $z \in Z$ ).

We can further define the notations in RBC based on the mode as follows.

The behavior of players in RBC is similar to that of chess, except that an additional sense step is added before the move step. Thus, each player's strategy in one turn contains two phases, sense and move.

In each turn, player  $i$  chooses actions (a sense action and a move action), by its strategy  $\sigma^i = (\sigma_s^i, \sigma_m^i)$ ,  $\sigma_s^i$  is player  $i$ 's sense strategy, and  $\sigma_m^i$  is move strategy.

Furthermore, action set  $A(h)$  in RBC is generated from the set of legal sense actions  $A_s(h)$  and the set of legal move actions  $A_m(h)$ ,  $A(h) = A_m(h) \times A_s(h)$ . Specifically, a sense action  $a_s = s \in A_s(h)$  locates a  $3 * 3$  area centered on  $s$  to be sensed.

**2.2.2. Information Set.** For IIGs, information set  $I_p$  for each player  $p$  is a partition of  $H_p$ . For any information set  $I \in I_p$ , any two states  $h, j \in I$  are indistinguishable to player  $p$ . Figure 2 uses RBC to give an example of information set.

In a game tree,  $I_p$  is a set of decision nodes of player  $p$ , which meets the following two conditions:

- (1) Each decision node in  $I_p$  is the decision node of player  $p$

- (2) When the game reaches a decision node in  $I_p$ , player  $p$  knows that it is in  $I_p$  but does not know which decision node of  $I_p$  it encounters

### 3. Heuristic Search of Uncertainty Control

A heuristic method is an approach for problem solving or self-discovery, which is not guaranteed to be optimal, perfect, or rational, but sufficient for reaching a feasible solution. While finding an optimal solution is impossible or impractical, heuristic methods can be used to speed up the process of finding a satisfactory solution. Heuristic search refers to a search strategy that attempts to optimize a problem by iteratively improving the solution based on a given heuristic function or a cost measure [11].

The heart of heuristic search methods is the idea of "continual researching" where a sound local search procedure is invoked whenever the agent must act without retaining any memory about how or why to reach the current state [12]. Our method for RBC game can be seen as a kind of heuristic search methods, using some measure function for better information exploitation. For solving RBC game, we divide the problem into two subproblems: how to control the explosive growth in scale of information sets and how to choose the most beneficial move action under imperfect information during the game. These two parts make up our heuristic search strategy.

**3.1. Heuristic Search for Sense Strategy in RBC Game.** The characteristic of RBC brings several difficulties to heuristic search in game solving. Firstly, since we cannot definitely know the opponent's knowledge and strategy, unreliable information may lead to an incorrect search direction. Secondly, how to control the growth of game states in information sets is another challenge. In order to ensure that the subsequent search is performed correctly, we must retain all possible states of the opponent as the current information set and cannot casually abandon any state. In addition, the player has a variety of action options. These two aspects have led to an information set of space rapid explosion and brought difficulty to storage.

Considering the above difficulties, we propose a novel sense method, HSUC, to prevent scale of real-time information set from exponential growth. Although RBC rules provide some extra information, such as notification of sense results, move results, and whether the player captures pieces, which can be employed to help reduce the uncertainty, HSUC is the major method to exploit all hints of RBC in our game system.

In this section, we will discuss how HSUC works to minimize the scale of the information set in RBC. An ideal way is to predict the opponent's next move action  $(s_f, s_t)$  and then take sense action  $a_{s_t}$  to get the sense result. A practical method for predicting opponents' actions is to absorb the idea of self-play, which is to use our own strategy to simulate the opponents' actions. Whenever we act sense action, we obtain the opponent's most likely action for every



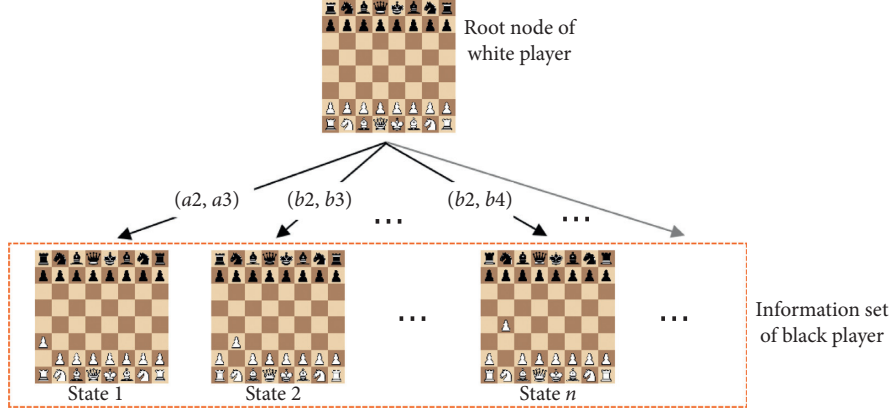


FIGURE 2: An example of the information set in RBC game. The root node is a white turn node. After the white player moves, the black player faces many indistinguishable possible states that are created by the opponent's unknown move; e.g., (a2, a3) and (b2, b3), are all possible move actions. These indistinguishable states in the dotted box form an information set of the black player.

remaining situation after the previous turn by move action selection strategy.

Unfortunately, the above approach is prone to bias. First of all, the initial premise of self-prediction is that the opponents adopt similar strategies to ours. When dealing with some specific agents, such as random or more powerful ones, sense actions will be severely misled which causes crashed performance of the whole system. Moreover, since the information set contains more than one state, plenty of sampling is required during the game tree search in order to guarantee the accuracy of state value evaluation.

HSUC focuses on estimating and reducing the whole uncertainty of the environment. Specifically in RBC, HSUC tries to find out the best sense square to minimize the number of possible states in real-time information set. Considering the game tree given in Figure 1, after the white player moves, the black player expands the game tree to form its real-time information set, which is the foundation of the next phase's strategy. To describe sense actions, the information set in the  $k$ -th turn is described as  $I_k: I_k = \{h_0, h_1, \dots, h_n\}$ ,  $|I_k| = n$  is the number of possible states in  $I_k$ . Each sense action reveals a  $3 \times 3$  sense square and helps eliminate some impossible states from  $I_k$ . For example, if the sense action reveals no piece in the sense square, all states with pieces in the sense square can be determined as "impossible states" and removed from  $I_k$ . Let  $A_s(I_k) = \{a_0, a_1, \dots, a_t\}$  and  $I_k(a_i)$  denote the reduced information set by taking sense action  $a_i$ . Then,  $I_k(a_i) \subseteq I_k$ ,  $m_i = |I_k(a_i)|$ . Employing  $g_k = (|I_k| - |I_{k+1}|) / |I_k| = 1 - (m_i/n)$  as the representative decay radio of sense action  $a_i$ , the target of heuristic search can be formalized as

$$a_i = \arg \max_{a_i \in A_s(I_k)} g_k = \arg \max_{a_i \in A_s(I_k)} \left(1 - \frac{m_i}{n}\right). \quad (1)$$

Here,  $g_k$  is employed to trace the tendency of game uncertainty. In this sense, the goal is to choose a sense action to maximize  $g_k$  for each turn of the game. However, there is no guarantee that the exact value of  $g_k$  can be found under the condition of imperfect information. So what should be done is to design a proper heuristic function  $H$  to evaluate  $g_k$ .

Firstly, we introduce how to evaluate the sense action's efficiency. Let  $D_i(h_1, h_2)$  indicate whether sense action  $a_i$  can distinguish between states  $h_1$  and  $h_2$  in  $I_k$ . As long as one of the 9 positions in the  $3 \times 3$  sense square is different (existence or types of pieces),  $D_i$  is set to 1; otherwise, it is set to 0. Figure 3 shows a specific example. The formula description is as follows:

$$D_i(h_1, h_2) = \max_{s_1 \in h_1(a_i), s_2 \in h_2(a_i)} L(s_1, s_2), \quad (2)$$

where  $s_1$  and  $s_2$  are the corresponding position in  $h_1$ 's and  $h_2$ 's sense squares of  $a_i$ .  $L(s_1, s_2)$  is defined as

$$L(s_1, s_2) = \begin{cases} 1, & P_{s_1} \neq P_{s_2}, \\ 0, & \text{otherwise,} \end{cases} \quad (3)$$

where  $P_{s_1}$  and  $P_{s_2}$  mean the pieces in positions  $s_1$  and  $s_2$ .

In this sense,  $D_i(h_1, h_2) = 0$  means states  $h_1$  and  $h_2$  cannot be distinguished by sense action  $a_i$ . Let  $H_{D_i}$  be the maximum subset of the current information set which contains the largest number of indistinguishable states given the sense action  $a_i$ .  $H_{D_i}$  satisfies the following three constraints:

- (1)  $H_{D_i} \subseteq I$ .
- (2)  $\forall h_1, h_2 \in H_{D_i}, D_i(h_1, h_2) = 0$ .
- (3)  $\forall h_i \in H_{D_i}, \sum_{h_j \in H_{D_i}} D_i(h_i, h_j) > 0$ .

We define that

$$C(a_i, I) = |H_{D_i}|. \quad (4)$$

Note that, in some cases, there may exist more than one maximum subset and the indistinguishable states they contain are different, but values of  $C(a_i, I)$  are the same. In this way, given sense action  $a_i$  and current real-time information set  $I_k$ , the sense efficiency can be described by the following heuristic function:

$$H(a_i, I_k) = n - C(a_i, I_k) = n - C(a_i, I). \quad (5)$$

```

Input: Information set of agent in turn  $k$ :  $I_k$ , The set of legal sense actions:  $A_s$ 
Output: Sense action:  $a_s$ 
(1) for each  $h_j$  in  $I_k$  do
(2)   Perform every legal move action for  $h_j$  to get successors
(3) end for
(4)   Combine all successors to update  $I_k$ 
(5)   Let  $T = 0, a_s = \text{None}$ 
(6) for each  $a_i$  in  $A_s$  do
(7)   for each two states  $(h_1, h_2)$  in  $I_k$  do
(8)     Calculate  $D_i(h_1, h_2)$  using formula (2)
(9)   end for
(10)  Find  $H_{D_i}$  and calculate  $C(a_i, I_k)$  using formula (4)
(11)  Calculate  $H(a_i, I_k)$  using formula (5)
(12)  if  $T < H(a_i, I_k)$  then
(13)     $a_s = a_i$ 
(14)     $T = H(a_i, I_k)$ 
(15)  end if
(16) end for
(17) return  $a_s$ .

```

ALGORITHM 1: HSUC algorithm.

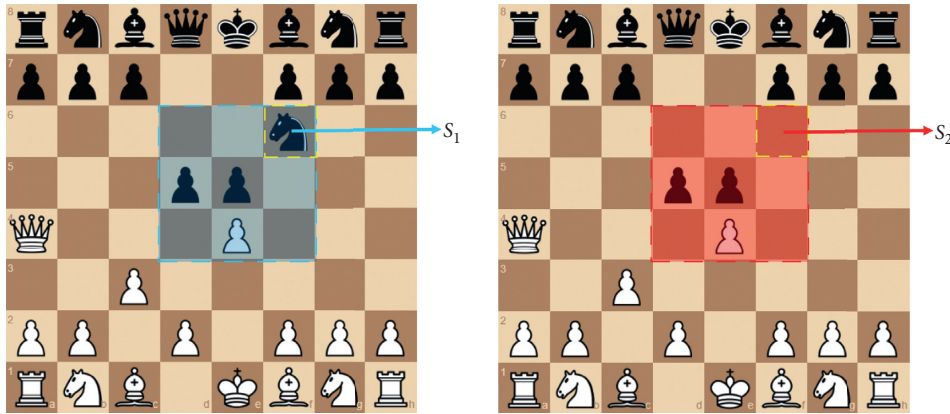


FIGURE 3: An example of evaluating sense efficiency. The two chess boards above denote two different states  $h_1$  and  $h_2$  in the current information set. As shown in the shaded part of the figure, when sense action  $a_i$  is adopted, a difference between the upper right corners  $s_1$  and  $s_2$  appear. In this case,  $L(s_1, s_2) = 1$  and  $D_i(h_1, h_2)$  is set to value 1.

By using the heuristic function  $H(a_i, I_k)$  to evaluate  $g_k$ , sense action can be searched in  $A_s(I_k)$  as follows:

$$a = \arg \max_{a_i \in A_s(I_k)} H(a_i, I_k). \quad (6)$$

By now, we have presented the details of HSUC method which adopts a heuristic searching approach to minimize the information sets' scale. The whole algorithm is shown in Algorithm 1.

**3.2. Foundation of the RBC Game System.** In this section, we will introduce the framework of our RBC game system incorporating the HSUC method. As shown in Figure 4, our architecture contains two main parts, HSUC for sense strategy and MCTS for move strategy.

When it is our turn in the RBC game, for example, at step  $t$ , firstly, we keep an information set of step  $(t-1)$  which contains all possible board states formed by our last move action. Then, we simulate all legal opponent's move actions on these board states to form the initial information set of step  $t$ . Generally, the scale of the information set will increase rapidly at the rate of dozens of times in this stage. And then, we apply our sense action provided by the HSUC algorithm. Powerful sense strategies will effectively eliminate impossible states as many as possible to get a reduced information set. At last, each remaining state will be solved as a root node by MCTS method and all of the returned solutions will be counted and the move strategy is determined by the statistics result. MCTS consists of four steps per iteration generally: selection, expansion, simulation, and backpropagation [13]. To control the iteration time, we use Stockfish to speed up

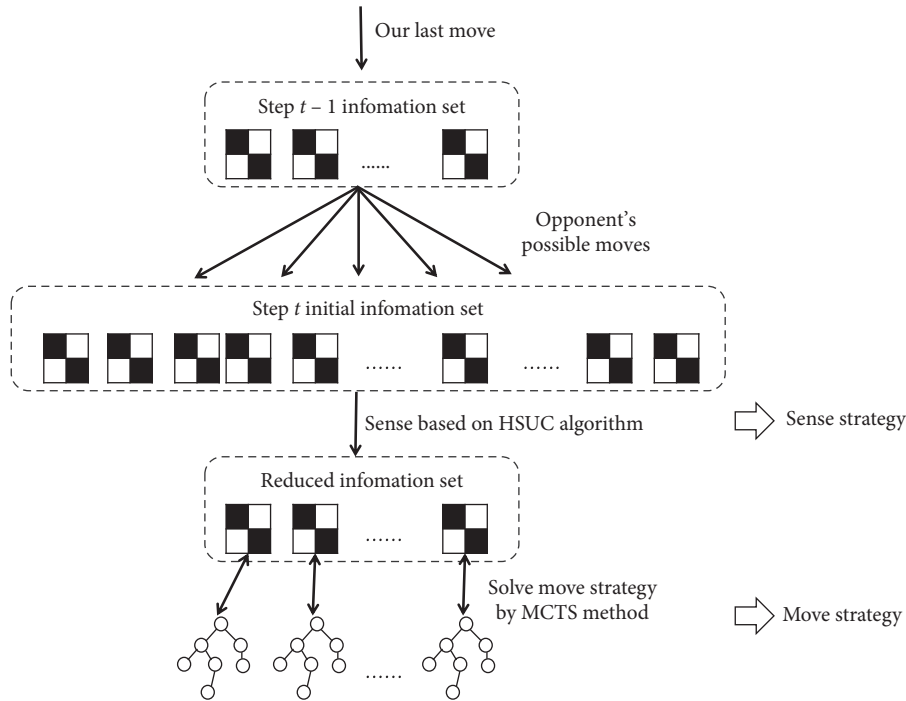


FIGURE 4: Architecture of our RBC game.

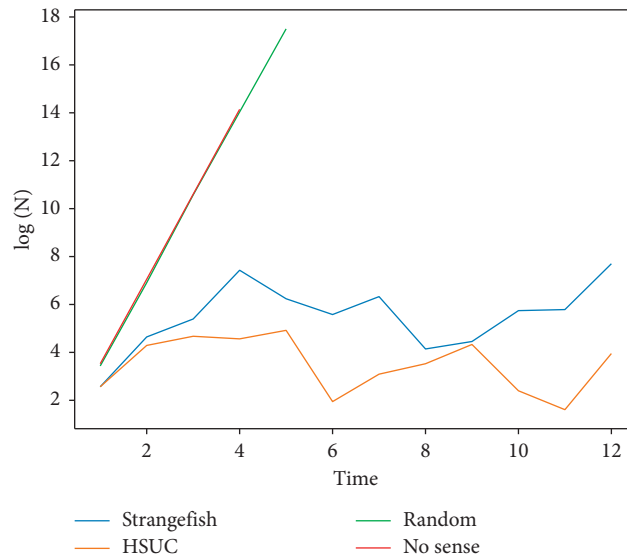


FIGURE 5: Result of different sense strategies.  $N$  is the number of states in the information set of RBC.

the termination of iterations and we constrain the depth of iterations based on the remaining time.

#### 4. Experiments

In this section, we evaluate the effectiveness of HSUC and the RBC game-solving framework mentioned in the previous section. During the experiment setting, we choose one agent from the NeurIPS 2019 tournament, Strangefish (<https://github.com/ginop/reconchess-strangefish>), as a comparison baseline. Strangefish ranks the first place in the tournament, which makes the comparison in the

experiments much more convincing. We conduct two experiments to verify the effectiveness of our proposed sense method HSUC (Section 4.1) and the performance of the overall RBC Game System (Section 4.2). The experiments are based on the package provided by the tournament's organizer, and the agents we implement all comply with the competition rules and restrictions.

*4.1. Performance of HSUC in RBC Sense Phase.* To illustrate the growth rate of the number of states in RBC's information set and the importance of a good sense method in solving

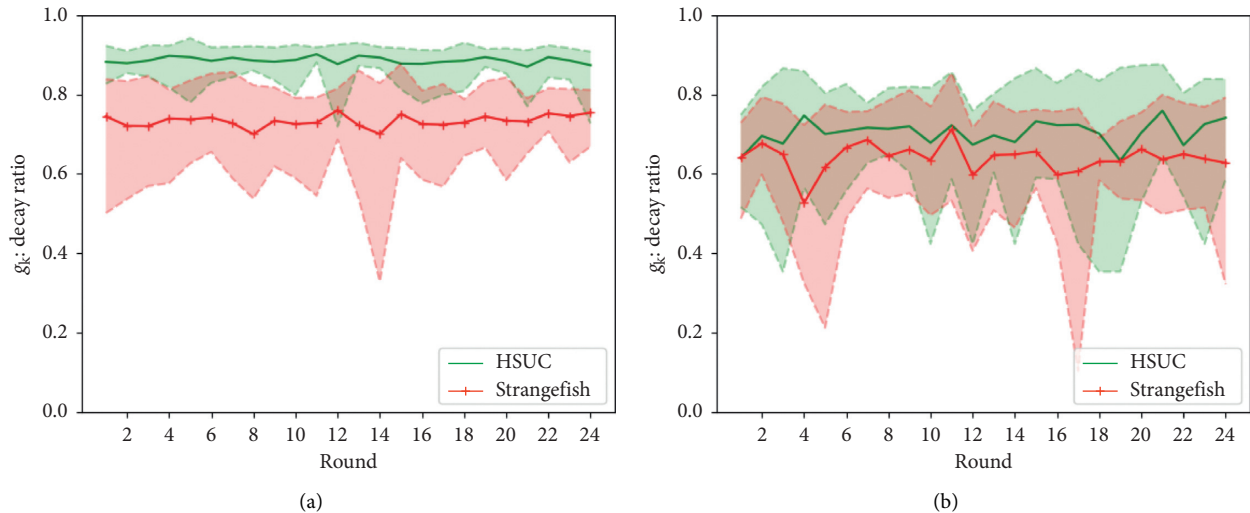


FIGURE 6: Efficiency of HSUC compared with sense method of Strangefish. Curves show the decay ratio  $g_k$  of the game during our agent with HSUC and Strangefish playing against different bots, respectively. The curve of our agent is higher than Strangefish in a large margin, which means that HSUC can maintain an obviously higher decay ratio of game uncertainty than Strangefish. (a) Against Random bot. (b) Against Trout bot.

IIGs like RBC, we conducted a comparative experiment of different sense methods firstly. In the experiment, agents with different sense methods play against the same opponent agent. During the game, the number of states in the real-time information sets after each sense action of each agent is tracked to obtain the experiment result.

As shown in Figure 5, the number of states in the information set of RBC increases exponentially without sense, and the problem of the exponential explosion of information sets can only be slightly alleviated with random sense actions. HSUC from our system and the sense method of Strangefish both perform better than the other two sense methods. We can conclude that the use of good sense methods can greatly reduce the scale of information set in RBC game.

The second experiment aims to verify the empirical advantages of HSUC. We let our RBC game system with HSUC compete with the baseline system at the platform of the NeurIPS 2019 tournament for 10 batches of games to obtain a statistical result. Each batch contains 24 rounds of games and each agent plays 12 rounds as black and 12 as white.

A robust sense method should satisfy the requests of efficiency and stability at the same time. First, decay ratio  $g_k$  mentioned in Section 3.1 is used to describe efficiency. The higher decay ratio denotes the method can reduce more impossible states by sensing. Second, the performance of the sense method should not fluctuate too much when facing opponents from different levels, which can be evaluated by the average scale of real-time information sets. The average scale of real-time information sets of turn  $j$  is  $N^j$ , which is calculated as follows:

$$\text{avg } N^j = \sum_{i=1}^T \frac{N_i^j}{T}, \quad (7)$$

where  $N_i^j$  denotes the average number of states of turn  $j$  in batch  $i$  and  $T$  is the total number of batches. We use the same experimental settings as above for the experiments.

The sense strategy of the Strangefish system is employed by scoring each move action to predict for sense area. The Strangefish method picks up the move with the highest score as the most likely action of the opponent and selects sense area based on the move action. It is similar to the method we introduce in Section 3.1.

The specific implementation of the experiment is to employ another agent as the opponent of our agent with HSUC and Strangefish, respectively, to collect data for calculating the indicator  $g_k$  for efficiency and  $\text{avg } N^j$  for stability during the game.

In Figure 6, we compare HSUC and the sense method of Strangefish by the decay ratio  $g_k$ . Figure 6(a) shows the result of playing against Random bot and Figure 6(b) is about the result of playing against Trout (another bot using Stockfish which performs better than random in the tournament). It can be seen that HSUC maintains a better performance against different bots than sense method of Strangefish. Moreover, HSUC shows more obvious advantages when playing against agents with a higher degree of randomness by reducing no less than 90% of uncertainty. The effective decay of uncertainty will bring great advantage to follow-up and can avoid the risk of failure due to timeout in game which is suffered by the agents maintaining a large number of states in the real-time information sets.

Figure 7 shows the average scale  $\text{avg } N^j$  of the real-time information sets of HSUC and Strangefish. The curves in the figure obviously indicate that our method performs better on managing scale of information sets than the method of the Strangefish against both of the rivals. In Figure 7(b), the maximal average scale of information sets of Strangefish even

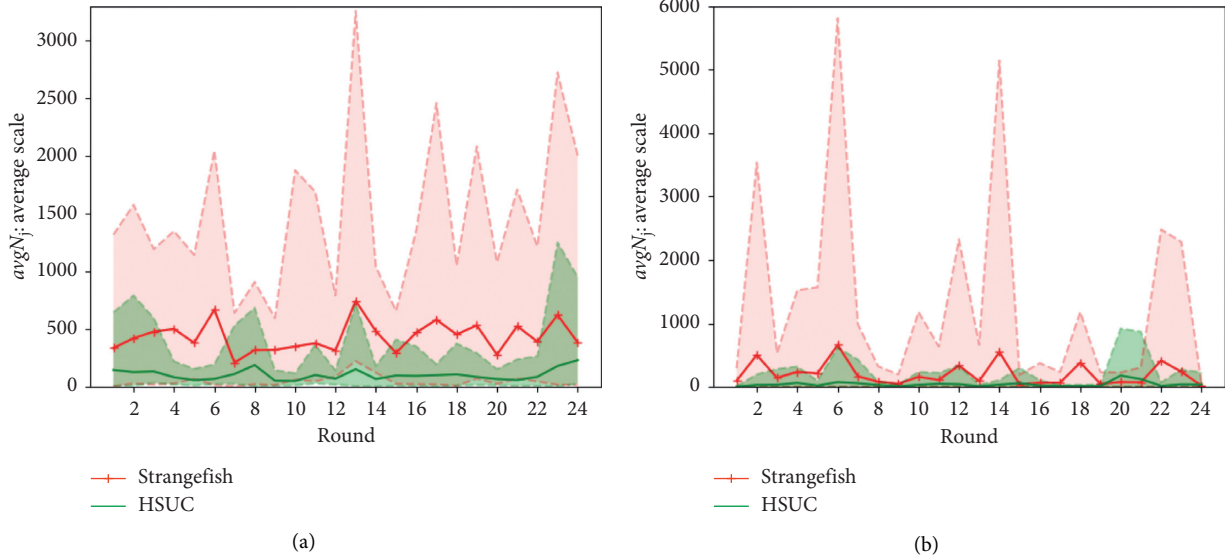


FIGURE 7: Stability of HSUC compared with the sense method of Strangefish. Curves show the average scale  $\text{avg } N^j$  of game during our agent and Strangefish playing against different bots, respectively. The curve of HSUC is lower and smoother than Strangefish plainly, which means that HSUC can control the scale of information sets more stably than Strangefish. (a) Against Random bot. (b) Against Trout bot.

TABLE 3: Uncertainty management rank.

| Bot name     | Institution         | Median state space size |
|--------------|---------------------|-------------------------|
| HSUC         | HIT (OUR)           | 12                      |
| Oracle       | JHU/APL             | 13                      |
| Wbernar5     | JHU                 | 18                      |
| La Salle Bot | La Salle University | 19                      |
| MBot         | SRC/Leela Chess Dev | 574                     |
| Random       | JHU/APL             | 50k+                    |

reached 6000 while that of HSUC is about 1000. The number of states of HSUC fluctuates smoothly while that of Strangefish fluctuates violently. Besides, considering Figure 6, we can conclude that the performance of HSUC is pretty stable for each turn and against different opponents from different levels.

By the way, the result of the NeurIPS 2019 tournament can also be a reference for the effectiveness of the sense method. As shown in Table 3, our sense method performs best on uncertainty management rank (<https://slideslive.com/38923177/reconnaissance-blind-chess-competition>).

**4.2. Performance of Our RBC Game System in NeurIPS 2019 Tournament.** In the NeurIPS 2019 tournament, each agent will fight against all the other opponents in turn by 24 rounds, and each agent begins with a cumulative 15-minute clock to make all their actions including sense and move. Our agent *A\_bot*, constructed with HSUC for sensing information and MCTS + UCT for move selection which incorporates new evaluation function Stockfish, achieves good result against many competitive opponents (such as agents from Microsoft and Google). For more details, please check here (<https://rbc.jhuapl.edu/tournaments/26>).

## 5. Conclusion

This paper introduces a novel method of uncertainty management in IIGs called HSUC. HSUC adopts a heuristic search process to guide sense actions to reduce the environment uncertainty of IIGs like RBC by minimizing the number of possible states in the real-time information sets. That is, HSUC can help agents to well understand the environment under imperfect information which enhances the effectiveness of game strategies. Furthermore, a viable RBC game system is realized by combining HSUC for sensing information and MCTS + UCT for selecting move actions. The experiments about HSUC and the RBC game system show that the scale of information sets is reduced effectively and efficiently through our method, providing convincing verification for the superiority of our method in terms of the uncertainty management in IIGs. In the future, we will conduct further research on factors affecting the uncertainty of the game environment and enrich the methods family of uncertainty management in IIGs.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this paper.

## Acknowledgments

The authors thank all the researchers in this field. This research was supported by Pingan-Hitsz Intelligence Finance Research Center, Key Technology Program of Shenzhen, China (no. JSGG20170823152809704), Key Technology

Program of Shenzhen, China (no. JSGG20170824163239586), and Basic Research Project of Shenzhen, China (no. JCYJ20180507183624136).

## References

- [1] J. Newman, C. L. Richardson, and S. M. Kain, "Reconnaissance blind multi-chess: an experimentation platform for ISR sensor fusion and resource management," in *Signal Processing, Sensor/Information Fusion, and Target Recognition XXV*, International Society for Optics and Photonics. SPIE, pp. 62–81, Bellingham, WA, USA, 2016.
- [2] R. Coulom, "Efficient selectivity and backup operators in monte-carlo tree search," in *Computers and Games*, pp. 72–83, Springer Berlin Heidelberg, Berlin, Germany, 2007.
- [3] M. Sustr, V. Kovarik, and V. Lisy, "Monte Carlo continual resolving for online strategy computation in imperfect information games," 2018, <https://arxiv.org/pdf/1812.07351.pdf>.
- [4] L. Kocsis and C. Szepesvári, "Bandit based monte-carlo planning," in *Machine Learning: ECML 2006*, pp. 282–293, Springer Berlin Heidelberg, Berlin, Germany, 2006.
- [5] D. E. Knuth and R. W. Moore, "An analysis of alpha-beta pruning," *Artificial Intelligence*, vol. 6, no. 4, pp. 293–326, 1975.
- [6] D. Silver, A. Huang, C. J. Maddison et al., "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.
- [7] N. Brown and T. Sandholm, "Superhuman ai for heads-up no-limit poker: libratus beats top professionals," *Science*, vol. 359, no. 6374, pp. 418–424, 2018.
- [8] N. Brown and T. Sandholm, "Superhuman AI for multiplayer poker," *Science*, vol. 365, no. 6456, pp. 885–890, 2019.
- [9] C. E. Shannon, "XXII. programming a computer for playing chess," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 41, no. 314, pp. 256–275, 1950.
- [10] J. Markowitz, R. W. Gardner, and A. J. Llorens, "On the complexity of reconnaissance blind chess," 2018, <https://arxiv.org/abs/1811.03119v2>.
- [11] J. J. Lu and M. Zhang, *Heuristic Search*, Springer, New York, NY, USA, 2013.
- [12] M. Moravčík, M. Schmid, N Burch et al., "Deepstack: expert-level artificial intelligence in heads-up no-limit poker," *Science (New York, N.Y.)*, vol. 356, no. 6337, pp. 508–513, 2017.
- [13] C. B. Browne, E. Powley, D. Whitehouse et al., "A survey of Monte Carlo tree search methods," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 4, no. 1, pp. 1–43, 2012.

## Research Article

# A Dynamic Hierarchical Clustering Data Gathering Algorithm Based on Multiple Criteria Decision Making for 3D Underwater Sensor Networks

Xiaoying Song<sup>1</sup>, Wei Sun<sup>1</sup> and Qilong Zhang<sup>2</sup>

<sup>1</sup>Department of Computer Science and Technology, Dalian Neusoft University of Information, Dalian 116023, China

<sup>2</sup>College of Computer Science and Engineering, Northeastern University, Shenyang 110819, China

Correspondence should be addressed to Qilong Zhang; zhangqilong@neusoft.edu.cn

Received 18 August 2020; Revised 20 September 2020; Accepted 6 October 2020; Published 22 October 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Xiaoying Song et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Data gathering is the basis of monitoring applications in an underwater sensor network, and excellent network coverage and data transmission reliability are the guarantees for the quality of monitoring tasks. However, the energy consumption of the nodes is too fast due to the heavy load of the cluster heads closer to the sink when data is transmitted between cluster heads (CHs) and the sink by multihop, which leads to an energy hole problem in an underwater sensor network of clustering technology. Aiming to address this problem, we propose a dynamic hierarchical clustering data gathering algorithm based on multiple criteria decision making (DHCDGA) in a 3D underwater sensor network. Firstly, the entire monitoring network is divided into many layers. For selecting a cluster head in each layer, multiple criteria decision making of an intuitionistic fuzzy Analytic Hierarchy Process (AHP) and hierarchical fuzzy integration is adopted. Furthermore, a sorting algorithm is used to form a clustering topology algorithm to solve the problem that there is the only node in one cluster. Then, an energy-balanced routing algorithm between clusters is proposed according to the residual energy of the node, the depth, and the number of neighbor nodes. Finally, the simulation results show that DHCDGA can not only effectively balance the energy consumption of the network and prolong the network lifetime but also improve network coverage and data gathering reliability.

## 1. Introduction

With the development of science and technology, the ocean, which covers 70% of the earth's surface area, plays an important role in the development of smart city [1, 2]. With the rise of the marine economy, countries all over the world pay more and more attention to the rights and interests of the ocean, and humans urgently need new technologies to observe and develop the ocean [3]. Underwater sensor networks are one of the hot ocean observation technologies that have emerged in the recent years [4, 5]. Underwater sensor networks have broad applications in many military and civilian fields, for example, for ocean exploration and development [6], for disaster warning and forecast [7], for environmental monitoring [8], and submarine detection [9]. Therefore, the

underwater sensor network has attracted widespread attention from domestic and foreign researchers.

Compared with traditional sensor networks [10], the main feature of underwater sensor networks is the use of underwater acoustic communication [11]. The absorption rate of radio signals in the water is very high, the signal propagation distance is limited, and the distance of underwater radio communication is generally not more than 100 meters. Therefore, it is not suitable for the needs of long-distance communication in underwater sensor networks [12]. Besides, the optical signal cannot be transmitted over long distances in seawater, its medium has a high absorption rate, and it is easily blocked, refracted, and reflected by various creatures and obstacles in the sea [13]. Therefore, underwater sensor networks use acoustic communication [14, 15]. The underwater acoustic channel has the

remarkable characteristics of high delay, high bit error rate, and low communication bandwidth [16, 17]. This makes the traditional sensor network communication protocol not directly applicable in underwater, which brings challenges to the protocol design of the underwater sensor network. In summary, the environment of underwater acoustic communication is harsh and the data transmission rate is very low. Therefore, it is of great significance to study the data gathering of underwater sensor networks.

There has been a lot of research studies on the data gathering strategy of underwater sensor networks. DBR [18] was a flat data gathering algorithm that provided a solution for the networking of underwater sensor networks in the early years. However, flat data gathering algorithms are often used in smaller networks. If the network scale is large, nodes close to the sink will undertake too much data forwarding and consume a lot of energy. This causes the nodes near the sink to die prematurely and form an energy hole. The death of these nodes will cause the rapid death of the surrounding nodes. Therefore, delaying the appearance of energy holes helps to prolong the network lifetime. Compared with the flat data gathering algorithm, the clustering data gathering algorithm can balance the energy load of the network. Hence, to adapt to a large-scale network environment, the use of clustering data gathering algorithms can alleviate the appearance of energy holes and effectively prolong the network lifetime.

At present, clustering data gathering algorithms are mainly focused on traditional ground wireless sensor networks. Its network topology is a two-dimensional flat network. These algorithms cannot be directly applied to the environment of a 3D underwater sensor network. Regarding the clustering data gathering algorithm of the 3D underwater sensor network, some theoretical research studies have been studied by many researchers. However, most of the algorithms are based on the clustering data gathering algorithm of the two-dimensional ground sensor network. These two-dimensional clustering algorithms have been transplanted into the three-dimensional underwater sensor network environment. They do not consider the mobility of nodes and unknown location information so that they cannot solve the problem of clustering data gathering in underwater sensor networks. The main deficiencies are as follows:

- (1) Little consideration is given to the impact of node mobility on the network. The clustering data gathering algorithms of two-dimensional sensor networks mostly use flat static sensor networks as the research premise, ignoring the impact of node movement on the network topology.
- (2) A two-dimensional sensor network can use GPS to determine the distance between nodes, but GPS cannot be used to locate an underwater sensor network. There is a large error in the way of wireless measurement distance, and underwater measurement distance has become another hot issue.
- (3) At present, most clustering data gathering of underwater sensor networks use energy as a measure to select cluster heads or design clustering algorithms

based on underwater characteristics. They rarely consider coverage and data transmission reliability. Hence, this is difficult to ensure network coverage and data collection reliability, which is a fatal defect for monitoring applications in underwater sensor networks. Accordingly, some scholars have begun to study cluster algorithms for sensor network coverage preservation and data gathering reliability. The data gathering algorithm with coverage preservation and reliability can reduce the rate of the decline of network coverage, at the same time, and the reliability of data gathering is guaranteed. Current research focuses on data gathering algorithm with the coverage preservation and data gathering reliability in traditional wireless sensor networks [19]. Little work on the clustering data gathering algorithm is suitable for coverage preservation and data gathering reliability in underwater sensor networks.

In this paper, we propose a dynamic hierarchical clustering data gathering algorithm based on multiple criteria decision making in 3D underwater sensor network. Aiming at these problems, the contributions of this paper are as follows:

The entire monitoring network is divided into layers. In each layer, multiple criteria decision making of an intuitionistic fuzzy AHP and hierarchical fuzzy integration is adopted in the phase of selecting a cluster head.

A sorting algorithm is used to form a clustering topology algorithm to solve the problem that there is only one node in one cluster.

According to the residual energy of the node, the depth, and the number of neighbor nodes, an energy-balanced routing algorithm is proposed between cluster heads.

In the simulation, DHCDGA is compared with leach-coverage-U and NULCPR, in terms of network lifetime and other five performances. The simulation results show that DHCDGA can not only effectively balance the energy consumption of the network and prolong the network lifetime but also improve network coverage and data gathering reliability.

The remainder of this paper is organized as follows. The related work of the field is introduced in Section 2. After introducing the network model and energy consumption model, in Section 3, a detailed description of the proposed clustering data gathering algorithm is presented in Section 4. In Section 5, we analyze the DHCDGA theoretically. Section 6 presents and analyzes six performance metrics of the DHCDGA in comparison with two other data gathering algorithms, and we conclude the paper in Section 7.

## 2. Related Work

By now, some work has been finished on improving energy efficiency or coverage preservation or data gathering reliability for 2D wireless sensor networks. However, there are



only a few clustering algorithms improving the above-mentioned three factors in a joint way for 3D underwater wireless sensor networks. Some clustering algorithms have been introduced in Section 1. In this section, we only briefly review some other typical clustering algorithms and considering coverage preservation and data gathering reliability.

Leach-coverage-U [20] was modified from the LEACH and the virtual grid routing protocols. The CPCHSA algorithm was added based on LEACH to achieve the best sensing coverage. Different nodes were assigned different probabilities of being a cluster head to maximize the network sensing coverage. However, Leach-coverage-U does not focus on data gathering reliability in underwater sensor networks. NULCPR [21] was a distributed unequal cluster size and hierarchical coverage-preserving routing algorithm. Spatially, the network was gradually built up from the sink node to the outside. Logically, a downward tree structure with the sink node as the root was constructed. The farther the node was from the Sink, the lower the level in this tree structure. Each layer ran the NCPR algorithm independently to complete the cluster head election, and each layer of the cluster head established a connection link with the upper node to ensure network connectivity. At the same time, the communication radius of nodes in each layer gradually increased as the number of network layers increased. In this way, the communication radius of the node was smaller and the cluster density was larger in the area close to Sink, and the communication radius was larger and the cluster density was smaller in the area far away from Sink. Hence, the energy consumption was balanced and the network lifetime is prolonged. However, NULCPR only considers energy efficiency and coverage preservation and ignores the reliability of data gathering.

CBEER [22] was an event-driven energy-efficient routing approach called clustering-based energy-efficient routing. In CBEER, a BS was responsible to run the routing scheme to optimize the cluster head (CH) selection for cluster creation. By cluster head selection, dynamic clusters were established. In this way, finding the shortest routing paths and consuming energy uniformly became possible in the entire network. Otherwise, a novel evolutionary algorithm based clustering was used to realize the reduction of energy consumption. For avoiding energy consumption of the same CHs, the CHs included in the current routing path were deleted from the other paths in the routing table. However, CBEER does not focus on coverage preservation and data gathering reliability. HENPC [23] was an energy-efficient clustering algorithm for magnetic induction-based underwater wireless sensor networks. The clustering protocol included two main parts. First, the jellyfish breathing algorithm was constructed based on the node contribution density to optimize the cluster size. Second, a node adjustment was built based on the Voronoi diagram to obtain an appropriate number of selected CHs. However, HENPC only focuses on energy efficiency, and it is not suitable for network scenarios with the request of high network coverage and data gathering reliability.

SMO [24] was a clustering approach based on spider monkey optimization. It resolved the node mobility caused

by the water current. Due to the node mobility, it led to transmission errors, link loss, collisions, and congestion, if not well handled. SMO was fit for heterogeneous underwater wireless sensor networks. There were three different roles in the network. They were cluster heads, local leaders, and global leaders. They worked together to provide energy-efficient data gathering. However, CBEER does not focus on coverage preservation and data gathering reliability. QERP [25] was a Quality-of-Service (QoS) aware evolutionary routing protocol for underwater wireless sensor networks. It addressed the challenge caused by reliable data delivery. For example, impairments of the acoustic transmission were caused by excessive noise, extremely long propagation delays, high bit error rate, low bandwidth capacity, multipath effects, and interference. QERP could improve the packet delivery ratio and reduce average end-to-end delay and overall network energy consumption. However, QERP is not suitable for network scenarios with the request for high network coverage. CUWSN [26] was an energy-efficient routing protocol selection for a cluster-based underwater wireless sensor network. In CUWSN, the multihops communication technique was adopted to reduce network energy. For further reducing energy consumption, CUWSN selected the cluster coordinator node and cluster head. The proposed CUWSN reduced power consumption by selecting the cluster coordinator node and cluster head to get a maximum lifetime of the network. However, CUWSN only considers energy efficiency, and it is not suitable for network scenarios with the request of high network coverage and data gathering reliability.

In [27], a fuzzy- and PSO-based clustering scheme using energy and distance parameters was proposed. The nodes were clustered by fuzzy clustering algorithms based on the geographical locations and the probability of belongingness of the sensor nodes. An improved PSO was used to select a cluster head. The process of cluster head selection considered multifactors to minimize energy consumption. However, the clustering scheme does not consider other factors, such as network coverage ratio and reliability of data gathering. MLCEE [28] was a multilayer cluster-based energy-efficient protocol for UWSNs. The goal of MLCEE mainly addressed the issue of a hotspot, high error rate, and high consumption of energy. In MLCEE, the entire network region was firstly divided into different layers from the surface to bottom and in every layer. Then, the second stage was the clustering of the nodes at the same layers. Finally, the cluster head selected the next-hop among the CHs based on the greater fitness value, small Hop-id, and small layer number. Although MLCEE has improved energy efficiency and data gathering reliability, it does not consider network coverage. FCMMFO [29] was a hybrid clustering method based on fuzzy c means and the moth-flame optimization method to improve the performance of the network. FCMMFO adopted FCM to divide the network. Theoretically, the optimal number of clusters could be determined by the elbow method. By MFO, the optimal locations of the cluster heads could be obtained. However, the clustering factor does not consider other factors, such as network coverage ratio and reliability of data gathering in an underwater sensor network.

### 3. System Model

In this section, we will introduce three models used in our clustering data gathering algorithm. They are the network model, the energy consumption model, and the node motion model.

*3.1. Network Model.* This paper assumes that there are the following properties in the underwater sensor network:

- (1) Sensor nodes are randomly deployed to form a three-dimensional underwater network, and the communication link between the nodes is reliable and two-way symmetrical.
- (2) There is a sink on the water. Its computing power and energy are not limited, and it can carry out wireless transmission.
- (3) The node moves within a certain range under the influence of water flow. We assume that the maximum offset distance is  $r$ .
- (4) Nodes can obtain their depth value, and each node is equipped with low-cost depth-sensing hardware.
- (5) The distance between nodes can be calculated according to the signal transmission strength.
- (6) DHCDGA is a periodic data gathering service. The nodes periodically gather data from sensors to sink in the underwater sensor network.

*3.2. Energy Consumption Model.* The algorithm in this paper adopts the same energy consumption model for underwater acoustic communication as in [30]. For underwater sensor nodes, the energy consumption of sending a message is about ten times that of receiving a message and idle listening. Therefore, the energy consumption of sending messages accounts for a large proportion of the total energy consumption, and reducing the energy consumption of sending messages means that the total energy consumption of the entire network can be reduced. The algorithm in this paper uses the energy consumption generated by sending the message as the main parameter to measure the total energy consumption of the entire network. Assume that  $P_o$  is the minimum power required by the node to receive messages normally; if the power attenuation function for the broadcasting distance  $x$  is  $A(x)$ , the transmission power should at least reach  $P_o A(x)$  to ensure that the node can receive the message. Assuming that the sending delay of a node sending  $l$  bit data is  $T_p$ , the energy consumed by sending  $l$  bit data  $E_{tx}(l, x)$  is

$$E_{tx}(l, x) = T_p P_o A(x), \quad (1)$$

in which  $A(x)$  is a function variable related to the underwater acoustic propagation model and the transmission frequency, which can be expressed as

$$A(x) = x^k a^k, \quad (2)$$

in which  $k$  is the related parameter of the underwater acoustic propagation model. When  $k$  is 1, it is a cylindrical

propagation model, and when  $k$  is 2, it is a spherical propagation model. Usually,  $k$  is 1.5 to represent the actual underwater acoustic propagation model.  $a$  is related to frequency  $f$  and can be obtained from the energy absorption coefficient  $\partial(f)$ .

$$a = 10^{\partial(f)/10}, \quad (3)$$

in which the energy absorption coefficient is

$$\partial(f) = 0.11 \frac{f^2}{1 + f^2} + 44 \frac{f^2}{4100 + f^2} + 2.75 \times 10^{-4} f^2 + 0.03. \quad (4)$$

*3.3. Node Motion Model.* To be able to simulate the motion state of the node in the water more accurately, the following conditions need to be considered when establishing the node motion model. Since the node is fixed at the bottom of the water, the depth of the node can be changed by adjusting the length of the anchor chain. The nodes in this state are affected by the water flow and the traction of the anchor chain and move within a certain range [31]. Figure 1 shows the node force in the underwater environment.

As shown in Figure 1, the force analysis of the node shows that the node is subjected to the lateral impact force  $F$ , buoyancy  $f$  of the water flow, and the tensile force  $T$  of the anchor chain to the node (ignoring the gravity of the node). These three forces constitute a set of balance forces. We set that the maximum angle between the anchor chain and the vertical direction is  $\alpha$ ,  $\tan \alpha = F/f$ . It is worth noting that we do not research on fluid mechanics. For simplification, the maximum offset distance of node movement is given in Section 6. Nodes perform random waypoint movement within this offset distance [32].

## 4. DHCDGA

Firstly, after the network is initialized, it will be layered according to the initial information. DHCDGA is implemented in each layer. The time unit of DHCDGA is round. Each round is divided into a cluster topology establishment phase and a data transmission phase. In the cluster establishment phase, multiple criteria decision making of an intuitionistic fuzzy AHP and hierarchical fuzzy integration is adopted to select cluster head so that cluster topology of unequal size is established. Namely, the cluster farther from the sink on the water surface has a larger cluster size, and the cluster closer to the sink has a smaller cluster size. In the data transmission phase, according to the remaining energy of the node, the depth, and the number of neighbor nodes, an energy-balanced routing transmission between cluster heads is proposed. The schedule architecture of DHCDGA is shown in Figure 2.

*4.1. Cluster Selection.* In this subsection, the cluster head nodes in the network are selected. Multiple criteria decision making of an intuitionistic fuzzy AHP and hierarchical fuzzy integration is adopted to select the cluster head. To achieve energy efficiency while taking into account service quality requirements, the main criteria for cluster head selection are

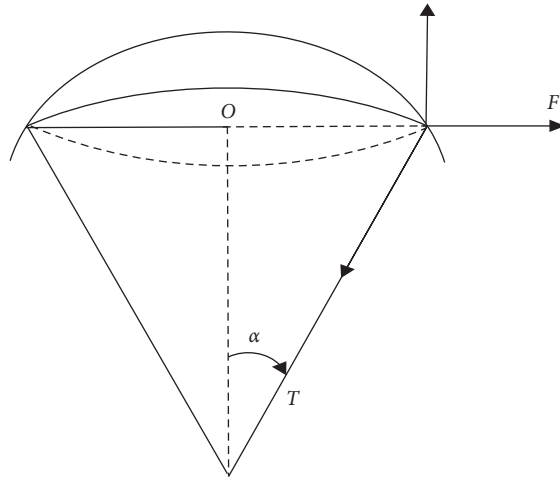


FIGURE 1: Node force.

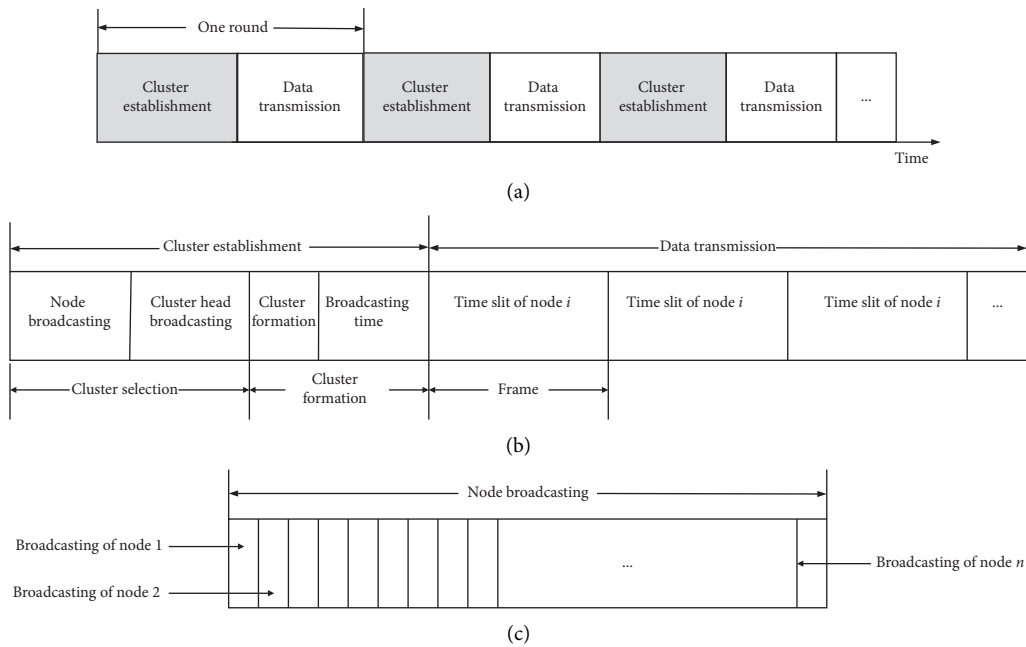


FIGURE 2: Schedule architecture of DHCDGA.

defined as energy status, service quality status, and node location status. Furthermore, each main criterion has two subcriteria. They are residual energy, message cost, coverage factor, link reliability, number of neighbor nodes, and depth factor. According to multiple factors affecting the selection of cluster heads, we have established a comprehensive evaluation hierarchical structure showed in Figure 3. It is mainly for a comprehensive evaluation of various indicators selected by cluster heads.

4.1.1. *Comprehensive Attribute Evaluation Value of the Cluster Head (h).* In the initialization phase, each node broadcasts an initialization message *INITIAL\_UN\_MES* to the network, which includes the node's ID, remaining energy, communication radius, and the distance between layers. The

node can locate the depth in the water to determine the level of the node. Nodes can estimate the distance between nodes based on the strength of their received signals. After the initialization phase is completed, each node can calculate the maximum and minimum residual energy of neighboring nodes, the number of neighboring nodes, and other information according to the neighbor node table. In this way, each node can obtain the evaluation value of each attribute in the comprehensive evaluation hierarchy of the cluster head.

(1) *Residual Energy.* In a clustered topology network, there is a huge difference in energy consumption between cluster heads and member nodes because the cluster head is responsible for receiving data from member node, fusing data, and relaying data to other cluster heads. In the underwater sensor network, the power supply cannot be replaced due to

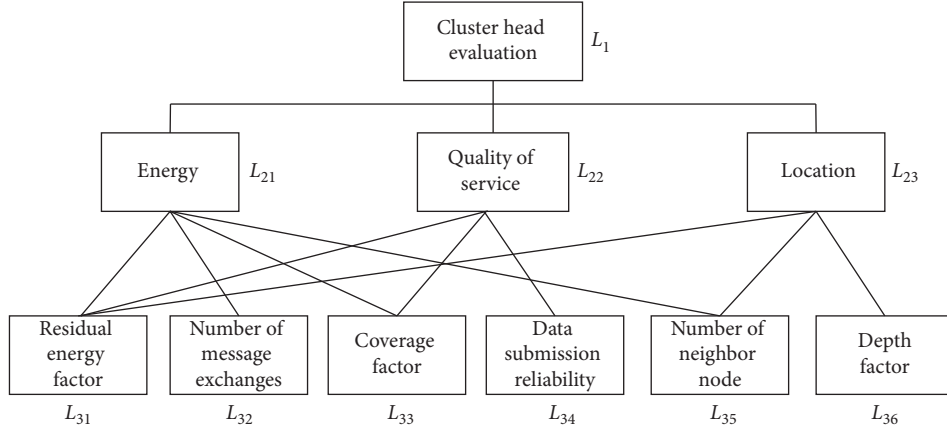


FIGURE 3: Comprehensive evaluation hierarchical structure of the cluster head.

environmental restrictions. Hence, energy is the most important and scarce resource. The evaluation value of residual energy is defined as follows:

$$E_i = \frac{E_r - E_{\min}}{E_{\max} - E_{\min}}, \quad (5)$$

in which  $E_r$  is the residual energy of node  $i$ .  $E_{\max}$  and  $E_{\min}$  are the maximum and minimum residual energy in the neighbor nodes of node  $i$ , respectively. The higher the value of  $E_i$  is, the more the residual energy is and the smaller the energy limit of the node is.

(2) *Number of Message Exchanges*. The number of message exchanges is mainly reflected in the total number of messages sent and received by nodes from the beginning of the network initialization stage to the establishment of the cluster topology. Since both sending and receiving messages consume the energy of the node, the number of exchanges of messages directly affects the energy consumption of the node. The evaluation value of the number of message exchanges is defined as follows:

$$M_i = \frac{M_T - M_C}{M_T}, \quad (6)$$

in which  $M_T$  is the total number of messages received and sent by all its neighbor nodes and  $M_C$  is the total number of messages received and sent by node  $i$ . The smaller the  $M_C$  value is, the higher the message cost  $M_i$  is.

(3) *Coverage Factor*. Coverage preservation is one of the most basic issues to ensure the Quality-of-Service (QoS). Selecting a node with better coverage as the cluster head can effectively prolong the functional time and make the residual energy of the node more. In a three-dimensional underwater sensor network, the coverage factor is the main factor to judge the QoS of a node. Its evaluation value is defined as follows [18]:

$$r(i) = \frac{\text{volume}\left(\left(\bigcup_{j \in N(i)} a_j\right) \cap (a_i)\right)}{\text{volume}(a_i)}, \quad (7)$$

in which  $a_i$  represents the perception area of node  $i$ .  $a_j$  represents the perception area of node  $j$ . Node  $j$  is the

neighbor node of node  $i$ .  $r(i)$  is the ratio of the volume intersection of the perception area of node  $i$  and its neighbor nodes to the perception area of node  $i$ . The larger the value of  $r(i)$  is, the larger the volume of intersection between node  $i$  and its neighbors is. Namely, the larger  $r(i)$  indicates that the node  $i$  is more likely to become the cluster head.

(4) *Data Transmission Reliability*. Data transmission reliability is an important issue to be considered for QoS because any node failure or packet loss will cause a large amount of packet loss in data transmission. In this way, more reliable requirements of data gathering are put forward for network scenarios with data collection reliability. The cluster head as the leader of the cluster unit is responsible for forwarding and receiving data more often in the network of clustered data gathering. Therefore, the probability of a highly reliable node as the cluster head should be greater. The evaluation value of data transmission reliability is defined as follows:

$$R_i = \frac{B_{\text{ava}}(i)}{B_{\text{total}}(i)}, \quad (8)$$

in which  $B_{\text{ava}}(i)$  is the available buffer space for node  $i$  and  $B_{\text{total}}(i)$  is expressed as the total buffer space of node  $i$ . We can observe that the smaller the available space of the node's buffer is, the lower the reliability of node  $i$  is.

(5) *Number of Neighbor Nodes*. In the network initialization, the number of current neighbor nodes can be determined according to the information in the neighbor information table. In an underwater sensor network where nodes are randomly deployed, the sink will calculate the optimal number of cluster heads in each layer when the energy is the smallest based on the global information of the nodes to obtain the optimal number of cluster members. According to this information, the closer the number of neighbor nodes is to the optimal number of cluster members, the greater is the probability that the node will become the cluster head. The evaluation value of the number of neighbor nodes is defined as follows:

$$N = \frac{|N_i - N_o|}{N_o}, \quad (9)$$

in which  $N_i$  is the number of neighbor nodes of node  $i$  and  $N_o$  is the best number of cluster members.

(6) *Depth Factor*. The depth of the node will affect the possibility of the node becoming the cluster head. The data gathering algorithm proposed in this paper is suitable for large-scale hierarchical underwater sensor networks. Since the selection of cluster heads is carried out in the same layer, the depth factor  $D(i)$  is calculated as the ratio of the relative height of the current node in the layer to the height of the layer. The depth factor is defined as follows:

$$D(i) = \frac{l \times L_{ni} - h_i + d \times \text{mod}(n-1, l/d)}{l}, \quad (10)$$

in which  $l$  is the height of the layer;  $h_i$  is the current depth information of the node  $i$ ;  $d$  is the downward adjustment distance of each network layer; and  $L_{ni}$  is the network layer of the  $n$ th round of node  $i$ . The calculation equation is as follows:

$$L_{ni} = \lfloor \frac{h_i + l - d \times \text{mod}(n-1, l/d)}{l} \rfloor. \quad (11)$$

*4.1.2. Determining the Attribute Importance Degree of the Cluster Head ( $g$ )*. In the previous section, the comprehensive attribute evaluation value of a cluster head has been obtained. This section determines that the attribute importance degree ( $g$ ) of the comprehensive evaluation of cluster heads will be determined according to the fuzzy integral. To be consistent with the actual selection of cluster heads in underwater sensor networks, it is necessary to use the intuitionistic fuzzy AHP that is closer to human thinking to select cluster heads because it can represent the neutrality of the expert's scoring. DHCDGA uses the cluster head selection comprehensive evaluation hierarchical structure in Figure 3 and the fuzzy analytic hierarchy process proposed in [33] to obtain the attribute importance degree ( $g$ ) of the cluster head. Firstly, many experts are asked to compare the importance of the indicators in the criterion level and the subcriteria for the upper-level target level. Based on the opinions of experts, the distance of intuitionistic fuzzy complementary judgment is established, where the intuitionistic fuzzy complementary judgment matrix of all attributes of the second layer to the target layer is shown as follows:

$$L_1 - L_2 = \begin{bmatrix} 0.5 & (0.7, 0.9) & (0.8, 0.9) \\ (0.1, 0.3) & 0.5 & (0.7, 0.8) \\ (0.1, 0.2) & (0.2, 0.3) & 0.5 \end{bmatrix}. \quad (12)$$

The intuitionistic fuzzy complementary judgment matrix of the attributes of the third-level subcriteria layer to the attributes of the second-level criterion layer is shown as follows:

$$L_{21} - L_3 = \begin{bmatrix} 0.5 & (0.8, 0.9) & (0.5, 0.8) & (0.8, 0.9) \\ (0.1, 0.2) & 0.5 & (0.1, 0.2) & (0.6, 0.7) \\ (0.2, 0.5) & (0.8, 0.9) & 0.5 & (0.7, 0.8) \\ (0.1, 0.2) & (0.3, 0.4) & (0.2, 0.3) & 0.5 \end{bmatrix}, \quad (13)$$

$$L_{22} - L_3 = \begin{bmatrix} 0.5 & (0.5, 0.8) & (0.3, 0.7) \\ (0.2, 0.5) & 0.5 & (0.5, 0.6) \\ (0.3, 0.7) & (0.4, 0.5) & 0.5 \end{bmatrix}, \quad (14)$$

$$L_{23} - L_3 = \begin{bmatrix} 0.5 & (0.3, 0.5) & (0.5, 0.7) \\ (0.5, 0.7) & 0.5 & (0.4, 0.7) \\ (0.3, 0.5) & (0.3, 0.6) & 0.5 \end{bmatrix}. \quad (15)$$

Furthermore, the fuzzy approximation judgment matrix of the abovementioned four matrices is obtained as follows:

$$F_{L1} = \begin{bmatrix} 0.5 & 0.875 & 0.889 \\ 0.125 & 0.5 & 0.778 \\ 0.111 & 0.222 & 0.5 \end{bmatrix}, \quad (16)$$

$$F_{L21} = \begin{bmatrix} 0.5 & 0.889 & 0.714 & 0.889 \\ 0.111 & 0.5 & 0.111 & 0.667 \\ 0.286 & 0.889 & 0.5 & 0.778 \\ 0.111 & 0.333 & 0.222 & 0.5 \end{bmatrix}, \quad (17)$$

$$F_{L22} = \begin{bmatrix} 0.5 & 0.714 & 0.5 \\ 0.286 & 0.5 & 0.556 \\ 0.5 & 0.444 & 0.5 \end{bmatrix}, \quad (18)$$

$$F_{L23} = \begin{bmatrix} 0.5 & 0.375 & 0.625 \\ 0.625 & 0.5 & 0.571 \\ 0.375 & 0.429 & 0.5 \end{bmatrix}. \quad (19)$$

Then, we realize Algorithm 1 in [33] by Matlab R2013 (b). The consistency test of the abovementioned four fuzzy complementary judgment matrices was performed, where  $\lambda = 0.5$ . After checking and adjusting the consistency of the matrix, a satisfactory consistency matrix is obtained as follows:

$$F'_{L1} = \begin{bmatrix} 0.5000 & 0.5521 & 0.5843 \\ 0.4479 & 0.5000 & 0.5464 \\ 0.4157 & 0.4536 & 0.5000 \end{bmatrix}, \quad (20)$$

$$F'_{L21} = \begin{bmatrix} 0.5000 & 0.8914 & 0.6512 & 0.9380 \\ 0.1086 & 0.5000 & 0.2466 & 0.5722 \\ 0.3488 & 0.7534 & 0.5000 & 0.7974 \\ 0.0620 & 0.4278 & 0.2026 & 0.5000 \end{bmatrix}, \quad (21)$$

$$F'_{L22} = \begin{bmatrix} 0.5000 & 0.5670 & 0.5132 \\ 0.4330 & 0.5000 & 0.4851 \\ 0.4868 & 0.5149 & 0.5000 \end{bmatrix}, \quad (22)$$

$$F'_{L23} = \begin{bmatrix} 0.5000 & 0.4516 & 0.5484 \\ 0.5484 & 0.5000 & 0.5710 \\ 0.4516 & 0.4290 & 0.5000 \end{bmatrix}. \quad (23)$$

Finally, the weight vector can be calculated by the (1) in [33]. After obtaining the satisfactory consistency matrix, the weight of all attributes in the second layer to the target layer can be obtained as  $W = (0.2879, 0.3352, 0.3769)$ . In the same way, the weights of the attributes subcriteria layer in the third level to the second level in the comprehensive hierarchy and the final combined weights of the attributes of the subcriteria layer in the third level relative to the elements of the target layer can be obtained. These weights are as shown in Table 1.

By the previous sections, the evaluation value ( $h$ ) of each attribute and its corresponding importance degree ( $g$ ) are obtained. Next, the overall evaluation value of the node can be obtained by the comprehensive evaluation hierarchical structure of the cluster head according to Figure 3. The process of calculating the overall evaluation value adopts the same tree structure as in [34]. Firstly, the attribute evaluation value is calculated from the leaf node. Then, the evaluation value of the previous layer can be obtained by the fuzzy integral to integrate the attribute evaluation value of the node and the weight of this layer. For instance, by the fuzzy integral, the evaluation value of the second layer L2 can be obtained by the evaluation value ( $h$ ) of the third layer L3 and the weight ( $g$ ) of the third layer. In the same way, by the fuzzy integral, the evaluation value of the first layer L1 can be obtained by the evaluation value ( $h$ ) of the second layer L2 and the weight ( $g$ ) of the second layer. By analogy, the evaluation value of the upper layer and the weight of this layer are calculated by fuzzy integral, and the comprehensive evaluation score of the node becoming the cluster head can be obtained.

## 4.2. Establishment of the Cluster Topology

**4.2.1. Cluster Head Broadcast Time.** In this section, the mapping relationship between the comprehensive evaluation value and cluster head broadcast time is illustrated. The cluster head broadcast timing method of the final cluster head adopts the method in [16]. Its core is to map the comprehensive evaluation score of the candidate cluster head to trigger the message of being a cluster head on the timeline. The time for publishing the competing cluster head message is shown in (24). Its purpose is to broadcast the message  $FIN\_CH\_MES$  for the cluster leader election quickly with high comprehensive evaluation scores.

$$T_i = \gamma \times (1 - CS_i) \times T_{ini}, \quad (24)$$

in which  $\gamma$  is a random number from 0.8 to 1;  $CS_i$  represents the comprehensive evaluation score calculated by the hierarchical fuzzy integral; and  $T_{ini}$  is the preset clustering time.

**4.2.2. Competition Radius of the Cluster Head.** After the cluster head is selected, the node needs to broadcast the

TABLE 1: Weights based on intuitionistic fuzzy AHP ( $g$ ).

| Attribute | $L_{21}$ | $L_{22}$ | $L_{23}$ | Weights |
|-----------|----------|----------|----------|---------|
|           | 0.2879   | 0.3352   | 0.3769   |         |
| $L_{31}$  | 0.0866   | 0.3066   | 0.3333   | 0.2533  |
| $L_{32}$  | 0.3454   | 0.0000   | 0.0000   | 0.0994  |
| $L_{33}$  | 0.1833   | 0.3606   | 0.2935   | 0.2843  |
| $L_{34}$  | 0.0000   | 0.3328   | 0.0000   | 0.1116  |
| $L_{35}$  | 0.3846   | 0.0000   | 0.0000   | 0.1107  |
| $L_{36}$  | 0.0000   | 0.0000   | 0.3731   | 0.1406  |

message in the range of the cluster head competition radius. Hence, clusters closer to the sink have the smaller cluster size, and clusters farther from the sink have the larger cluster size. The specific regulations are as follows.

We set that the network layer  $i$  starts from 0 and is numbered with increasing from the top to bottom. In the initial stage of the network, the NO.0 layer does not exist, and the network number starts from 1. The network number starts from 1. After that, the 0th layer of the network is within the communication range of Sink, so the 0th layer is required to not cluster and send data directly to Sink. The cluster head at other levels needs determination of its level according to (10), and then, the cluster head competition radius is determined by the following equation:

$$CH(i).cr = CH(i).br + level(j) \times initial\_range, \quad (25)$$

in which  $CH(i).cr$  is the competition radius of the cluster head  $i$ ;  $CH(i).br$  is the broadcast radius of the cluster head  $i$ ;  $level(j)$  is the number of layers; and  $initial\_range$  is the initial range value. From (25), it can be seen that the larger the cluster head is, the larger the competition radius is.

**4.2.3. Cluster Topology Formation.** In cluster topology formation, if a node loses the opportunity to become the cluster head, it will join a cluster as a common member node. In joining a cluster, there are two very important parameters, namely, the distance from the node to the cluster head and the density of the cluster head. If there is more than one cluster head in the communication range of a node, the node will calculate the score of joining each cluster head node according to the following equation:

$$Sort_{CH} = \alpha \cdot distance_{itoCH} + \beta \cdot Density_{CH}, \quad (\alpha + \beta = 1). \quad (26)$$

The node's decision to join which cluster head mainly depends on its lower score. Then, this node sends a message  $JOIN\_CH\_MES$  to the cluster head with the lowest score to join it. The cluster topology formation of other rounds is similar to this process. Algorithm 1 shows the pseudocode of the member nodes joining the cluster unit.

**4.3. Data Transmission.** After establishing cluster topology, the nodes begin to transmit data steadily. When the energy value of a node is lower than the threshold  $E_0$ , it means that the node is dead, and then, the network starts to reconstruct.

Input: the distance between node  $i$  to CH; the density of CH;  
Output: cluster topology

- (1) **for** each node  $i$  ( $i$  is not CH)
- (2) Receive message with the distance between node  $i$  to CH; the density of CH
- (3) Calculate  $\text{Sort}_{\text{CH}} = \alpha \cdot \text{distance}_{i\text{toCH}} + \beta \cdot \text{Density}_{\text{CH}}$
- (4) Select Min ( $\text{Sort}_{\text{CH}}$ )
- (5) Send request message  $\text{JOIN\_CH\_MES}$  to the CH which has the minimum score.
- (6) **end for**

ALGORITHM 1: Cluster topology formation.

In the data transmission phase, each cluster member node periodically gathers data from the sensing area and transmits the data to its cluster head according to TDMA timing allocated by the cluster head to avoid message collision in the same cluster. When all the data of the cluster members are transmitted to the cluster head, the cluster head will fuse the data and transmit the fused data packets to Sink. Therefore, data transmission can be divided into two stages: intracluster communication and intercluster communication. Cluster members sense the data in the gathering environment and transfer the gathered data to their cluster head. This process is intracluster communication. Direct transmission is adopted to simplify the communication between member nodes and cluster heads. In intercluster transmission, a multihop inter layer routing algorithm is proposed. The cluster head with a large layer number is transmitted to the smaller layer number and, finally, to Sink. This can not only ensure the direction of data transmission but also prevent the generation of routing loops. Before selecting the next-hop node, the cluster head or node  $i$  calculates the set of neighbor nodes in the upper layer according to the location information of the sink and establishes a table of neighbor nodes, including node ID, node depth, sink location, current remaining energy, and the number of neighbor nodes. The cluster head  $i$  calculates the routing cost of its upper neighbor nodes and selects the next-hop node of the cluster head with the lowest routing cost. When the cluster head  $j$  is selected as the next-hop node, the relay node of the next-hop is selected according to the depth information. The cost function is shown as follows:

$$\text{cost}(i, j) = \alpha \frac{d_{j\text{-Sink}}}{d_{i\text{-Sink}}} + \beta \frac{\bar{E}_{\text{above-Nei}}(i)}{E_{\text{residual}}(\text{CH}_j)} + \gamma \frac{\bar{N}_{\text{above-Nei}}(i)}{N(\text{CH}_j)}, \quad (27)$$

in which  $d_{i\text{-Sink}}$  is represented as the depth information of the node  $i$ ;  $d_{j\text{-Sink}}$  is the depth information of the upper neighbor node  $j$ ;  $\bar{E}_{\text{above-Nei}}(i)$  is the average residual energy of the neighbor nodes in the upper layer of the node  $i$ ;  $E_{\text{residual}}(N_j)$  is the current residual energy of the cluster head  $j$ ;  $\bar{N}_{\text{above-Nei}}(i)$  denotes the average number of neighbor nodes in the upper layer of the node  $i$ ; and  $N(\text{CH}_j)$  is the number of neighbors of the node  $j$ .  $\alpha$ ,  $\beta$ , and  $\gamma$  are weighted coefficients, and the sum of the three weighting coefficients is 1. The cluster head or node  $i$  selects the node whose cost function is the least among its neighbors, namely,  $k = \text{armincost}(i, j)$ .

When designing this cost function, a relay node closer to the sink may be selected in the first term of (27). Since the transmission range of the node is fixed in this paper, choosing a relay node closer to the sink can reduce the number of relays and reduce the energy consumption of the node.

In the second term, the node with higher remaining energy may be selected as the next-hop relay node. Because the energy of the underwater sensor network is limited, it consumes energy during transmitting data, receiving data, or selecting cluster heads. Therefore, the residual energy of the node is the main factor in selecting the relay node.

In the third term, the node with more neighbor nodes may be selected as the next-hop relay node. This factor is considered because the nodes are randomly deployed in this paper, namely, some nodes are deployed densely and some nodes are deployed sparsely. When the nodes are dense, the selected next-hop relay nodes are more scattered, and the energy consumption is more balanced. When the nodes are sparse, the selected next-hop relay nodes are more concentrated. This brings the nodes with the largest energy consumption to appear prematurely, which shortens the network lifetime.

It is worth noting that if there are multiple same minimum routing costs in the neighbor node table toward the water surface, a node is selected randomly as the next-hop relay node. If the set of neighbor nodes of the upper level of the cluster head  $i$  is empty, the fallback mechanism in [14] is adopted. Namely, this node is deleted from the neighbor information table, and the node with the second smallest cost function in the upper-level neighbor node set of  $\text{CH}_i$  is reselected as the next-hop relay node.

## 5. DHCDGA Theoretical Analysis

**Proposition 1.** *The message complexity of DHCDGA is  $O(N)$  in the clustering phase, where  $N$  is the number of sensor nodes in the network.*

*Proof.* Because the energy consumption of sending messages is much greater than that of receiving messages, the complexity of sending messages as the message complexity of DHCDGA is discussed only in the clustering phase. Firstly, in the initialization phase,  $N$  sensor nodes broadcast initialization messages and establish the neighbor table of the nodes. Furthermore, DHCDGA is executed in each layer

to form clustering topology. Because three communications are required between nodes, the total number of messages sent is  $N + N + N$ . Otherwise, assuming that  $M$  nodes in the network become the final cluster heads,  $M$  nodes broadcast, at least,  $FIN\_CH\_MES$  messages. There are  $(N - M)$  non-cluster heads, and they broadcast  $(N - M)$   $JOIN\_CH\_MES$  messages to the cluster head with the smallest score to construct a clustering topology. In summary, the total message overhead is  $4N + M + (N - M)$ . Therefore, the message complexity is  $O(N)$  in the clustering phase.  $\square$

**Proposition 2.** *In DHCDGA,  $(2K + l/d_1 - 1)$  hops are required, at most, for any data packet transmitted to Sink, where  $K = \lfloor D/l \rfloor$ ,  $D$  is the area radius of the underwater sensor network;  $l$  is the width of the layer; and  $d_1$  is the average distance of a hop in the first layer.*

*Proof.* Assuming that the node  $S_i$  is located in the  $K$ th layer, the node in its cluster has one cluster head. There are two transmission methods for a data packet transmitted from the  $K$ th layer to the Sink. Namely, the second layer to the  $K$ th layer is clustered, and the flat multihop transmission mode is used in the first layer.

Firstly, the number of transmission hops is calculated from the  $K$ th layer to the second layer. The data packet of node  $S_i$  is transmitted to its cluster head, and the cluster head is transmitted to the cluster head in the  $(K - 1)$  layer after data fusion. Keeping this way, data packets are transmitted until the second layer. Therefore, the number of hops for data packets of  $(2K - 1)$  nodes is  $(2K - 2)$  in data transmission.

Secondly, the number of hops is calculated for flat multihop transmission in the first layer. When data packets are transmitted from the second layer to the first layer, different transmission methods will be used. In the first layer, the flat multihop data transmission is used, and the number of hops for the data packet transmitted to the sink is  $l/d_1$ . Another hop is from CH in the second level to the nodes in the first layer.

Accordingly,  $(2K + l/d_1 - 1)$  hops are required, at most, for any data packet transmitted to the sink in DHCDGA.  $\square$

**Proposition 3.** *When the cluster head transmits data to Sink, there is no routing loop.*

*Proof.* Proof by contradiction is adopted for Property 3. We assume that there is a routing loop  $\{CH_1, CH_2, \dots, CH_K, CH_1\}$ .  $D_i$  represents the depth of  $CH_i$ . According to the network model in this paper, the depth  $D_{i-1}$  of  $CH_{i-1}$  is shorter than  $D_i$ , namely,  $D_1 < D_2 < \dots < D_L < D_1$ . However, the depth  $D_L$  of the node  $CH_L$  is longer than the depth  $D_1$  of the cluster head  $CH_1$ . This contradicts the actual situation. Accordingly, it is guaranteed that no routing loop will occur when the cluster head transmits data to Sink.  $\square$

## 6. Simulation

**6.1. Simulation Scenarios and Parameter Setting.** The simulation environment is the Intel Pentium dual-core processor (2.2 GHz), 2G memory, and the experimental platform is Matlab R2013 (b). To analyze the effectiveness of DHCDGA, three clustering data gathering algorithms, DHCDGA, NULCPR, and LEACH-Coverage-U, are implemented.

We analyze and compare the six performances of three clustering data gathering algorithms whose network lifetime, the number of alive nodes, the average remaining energy, the number of cluster heads, coverage ratio of the network, and the total amount of data are received by Sink. In the simulation, 1000 sensor nodes are randomly deployed in  $100\text{ m} \times 100\text{ m} \times 100\text{ m}$  of the underwater three-dimensional monitoring area  $A$ . The coordinates of the sink are (50, 50, 100) (m). For calculating the coverage ratio, area  $A$  is divided into  $100 \times 100 \times 100$  small cubes, and the size of each cube is  $1\text{ m} \times 1\text{ m} \times 1\text{ m}$ . The simulation result is the average of twenty experiments, and other parameters are shown in Table 2.

**6.2. Network Lifetime.** The network lifetime of three data gathering algorithms is shown in Figure 4. FDT (First node Died Time) is the time when the node is dead first in the network. LDT (Last node Died Time) is the time when the last node is dead in the network. FDT is one of the most important indicators investigated by underwater sensor network data gathering algorithms. However, when the first node died in the network, if the node has strong coverage, it will not have a great impact on the data gathering of the whole network. Therefore, it is not enough to observe only FDT but also LDT of three clustering algorithms.

As shown in Figure 4, we can observe that the FDT of the Leach-coverage-U algorithm appears the earliest, while the LDT appears at the latest. This result can be explained by the fact that Leach-coverage-U transmits data from CH to the sink by a single hop in the data transmission phase. The cluster head consumes more energy.

Otherwise, in Leach-coverage-U, the cluster structure must be rebuilt in each round. The message is expensive and consumes a lot of energy. Therefore, the FDT of the lease-coverage-U algorithm appears the earliest. Also, single-hop data transmission will bring that the nodes far away from the sink are not able to transmit data to the sink due to insufficient transmission range until their energy is exhausted. Hence, the LDT of Leach-coverage-U appears the latest. NULCPR and DHCDGA are unequal clustering data gathering algorithms. The FDT of DHCDGA is about 18.71% longer than that of NULCPR. The LDT of DHCDGA is about 15.38% longer than that of NULCPR. This result can be explained by the fact that the multihop routing path constructed takes into account factors such as residual energy, depth, and the number of neighbor nodes during the data transmission. But, NULCPR only focuses on the distance between cluster heads.



TABLE 2: Simulation parameters.

| Parameters                              | Value       |
|---|-------------|
| Node throughput                         | 0.25 (kB/s) |
| The initial energy of the node          | 20 J        |
| Lowest receiving power ( $P_o$ )        | 3 (mW)      |
| Size of a data packet ( $l$ )           | 4000 (bits) |
| Communication radius ( $R$ )            | 600 (m)     |
| Sense radius                            | 15 (m)      |
| Carrier frequency ( $f$ )               | 10 (kHz)    |
| Energy diffusion                        | 1.5 (K)     |
| Energy consumption of data fusion       | 5 (nJ/bit)  |
| Maximum offset distance of node ( $r$ ) | 10 (m)      |
| Initial radius                          | 10 (m)      |

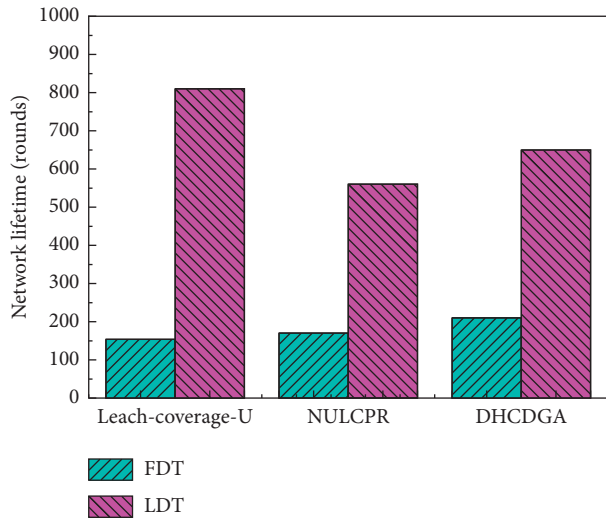


FIGURE 4: Network lifetime.

6.3. *Number of Alive Nodes.* The number of alive nodes represents the change in the number of nodes with energy from the beginning of work to the exhaustion of energy. This indicator is to examine the stability of the data gathering algorithm. The number of alive nodes for the three data gathering algorithms is shown in Figure 5.

As shown in Figure 5, the stability of DHCDGA is slightly better than that of Leach-coverage-U and NULCPR. Although DHCDGA occurs FDT, the number of alive nodes decreases linearly until all the nodes die instead of the network split causing the data gathering function paralysis. Since the message timing mechanism adopted by both DHCDGA and NULCPR replaces the message negotiation mechanism of Leach-coverage-U, energy consumption in the clustering phase is less than Leach-coverage-U. Otherwise, NULCPR only focuses on the factor of distance in selecting CH and data transmission; however, DHCDGA integrates multiple attributes for cluster head selection and routing path construction so that it is more suitable for underwater sensor networks with nonuniform node distribution. In summary, DHCDGA has better stability than the other two data gathering algorithms.

6.4. *Average Residual Energy of the Network.* The average residual energy of the network is the ratio of the current

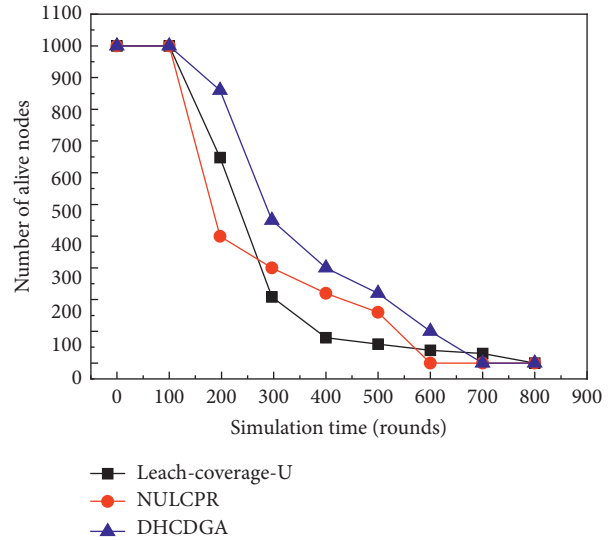


FIGURE 5: Number of alive nodes.

residual energy of each node in each round to the total number of nodes, which mainly examines the balance of energy consumption. The average residual energy of the three data gathering algorithms is shown in Figure 6.

As shown in Figure 6, DHCDGA consumes less energy than NULCPR. This result can be explained by the fact that DHCDGA considers six factors in CH selection and the competition radius has two factors when network forms cluster topology. These are results in the cluster structure more uniform and reasonable when nodes join in cluster units. Meanwhile, the intermultihop route algorithm takes into account factors such as the residual energy, depth information, and the number of neighbor nodes in the interdata transmission, which makes the selection of the next-hop relay node more reasonable and reduces energy consumption.

Otherwise, the nodes of Leach-coverage-U that is an equal size cluster algorithm have a less average remaining energy because the equally clustered data gathering algorithm does not use an effective balance mechanism for reducing energy consumption. The data gathering algorithm with equal cluster size may produce clusters with only one CH and no member nodes, which leads to an unbalanced node load and huge differences in energy consumption. Besides, Leach-coverage-U may cause CHs to be concentrated in certain areas. More edge nodes are generated, which is a result when all CHs are not within their communication range. Long-distance communication is required between edge nodes and sink, which increases energy consumption.

6.5. *Number of Cluster Heads.* The number of cluster heads is the total number of cluster heads in the network in different simulation times. This indicator is mainly to investigate the stability of the clustering algorithm in the data gathering algorithm. The number of cluster heads with changes in network time is shown in Figure 7.

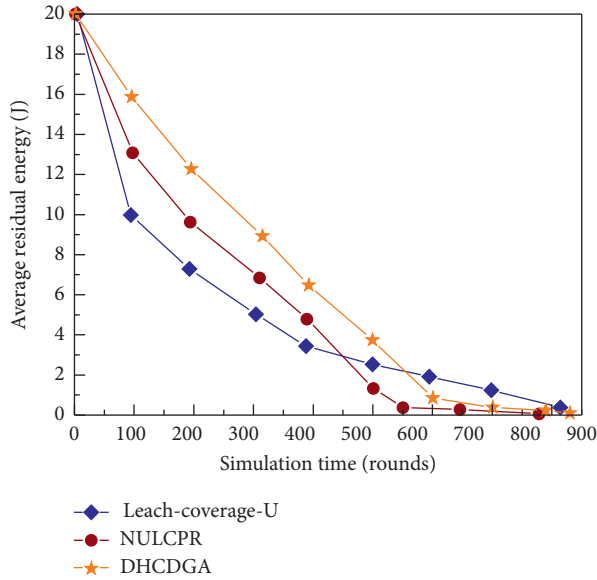


FIGURE 6: Average residual energy.

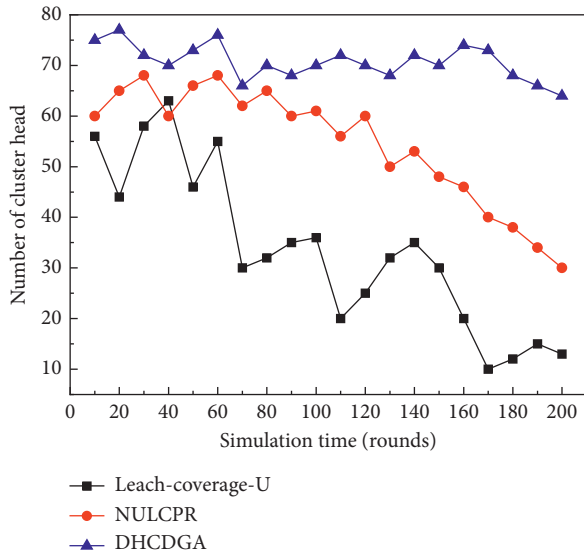


FIGURE 7: Number of cluster heads.

As shown in Figure 7, compared with NULCPR and Leach-coverage-U, the number of CH for DHCDGA is relatively stable. Because the number of CHs in the DHCDGA mainly depends on the number of neighbor nodes of  $CH_i$ 's every layer, network coverage, and other information, the cluster head competition radius of each layer is relatively fixed.

Therefore, the number of cluster heads in the DHCDGA does not change significantly, so its clustering algorithm has a better stability. NULCPR algorithm is the data gathering algorithm with coverage preservation. After around, there are no nodes to die, so the network topology has not changed so that the coverage redundancy of each node has not changed. NULCPR can continue to keep the cluster topology of the previous round, and the number of CH is relatively

stable. However, the number of CHs fluctuates sharply for Leach-coverage-U since Leach-coverage-U is a random clustering data collection algorithm. Cluster topology must be rebuilt in each round, and candidate CHs will be randomly assigned according to the proportion of alive nodes, which will generate a cluster with a single node. Accordingly, the number of CHs in Leach-coverage-U fluctuates sharply.

**6.6. Network Coverage Ratio.** Network coverage refers to the percentage of the total area of the node gathering range to the entire network area. The network coverage ratio for three data gathering algorithms is shown in Figure 8. Because the coverage ratio is closely related to the number of alive nodes, Figures 5 and 8 are observed together.

As shown in Figures 5 and 8, comparing with NULCPR and Leach-coverage-U, DHCDGA has certain advantages. Furthermore, the coverage ratio decreases approximately linearly after occurring FDT of three data gathering algorithms. Besides, there is no network fragmentation caused by the simultaneous death of a large number of nodes. The coverage rate of the NULCPR algorithm is higher than that of Leach-coverage-U.

As shown in Figure 5, Leach-coverage-U has more alive nodes than NULCPR before the network runs 265 rounds. Leach-coverage-U and NULCPR have the same alive nodes in the 265th round. As shown in Figure 8, in the 265th round, although Leach-coverage-U has more alive nodes at this time, the coverage ratio of NULCPR is about 8.62% higher than the Leach-coverage-U. At the same time, the coverage ratio of DHCDGA is about 39% higher than that of the Leach-coverage-U. Assuming that the expected coverage ratio of the monitoring area is about 70%, the number of running rounds that meet the requirements for Leach-coverage-U, NULCPR, and DHCDGA is about 212 rounds, 253 rounds, and 384 rounds, respectively. Compared with Leach-coverage-U, the network lifetime of NULCPR and DHCDGA has been prolonged by about 41 and 172 rounds, respectively.

**6.7. Total Amount of Data Received by Sink.** The total amount of data received by the sink refers to the total amount of data received by the sink in each round. The total amount of data received by the sink for three algorithms is shown in Figure 9.

As shown in Figure 9, The data packets received by the sink for three data gathering algorithms increase steadily as the running time increases. Near the end of the simulation, the slope of the curve gradually decreases, which is caused by the continuous death of nodes as time increases.

Besides, the total amount of data received by the sink for DHCDGA is more than that received by the other two algorithms, since DHCDGA considers the reliability of data gathering in CHs selection. Otherwise, neither NULCPR nor Leach-coverage-U considers the reliability of data collection, and NULCPR receives more data than Leach-coverage-U. Because NULCPR is a data gathering algorithm with unequal cluster size, Leach-coverage-U is a data gathering algorithm with equal cluster size. The network lifetime of an unequal clustering algorithm is longer than that of an equal

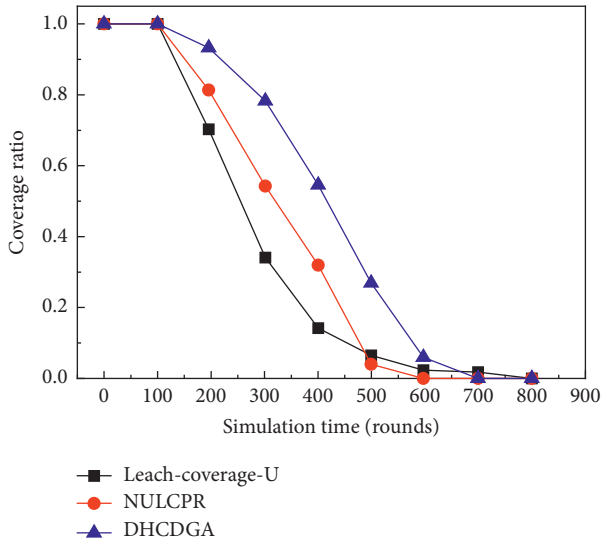


FIGURE 8: Network coverage ratio.

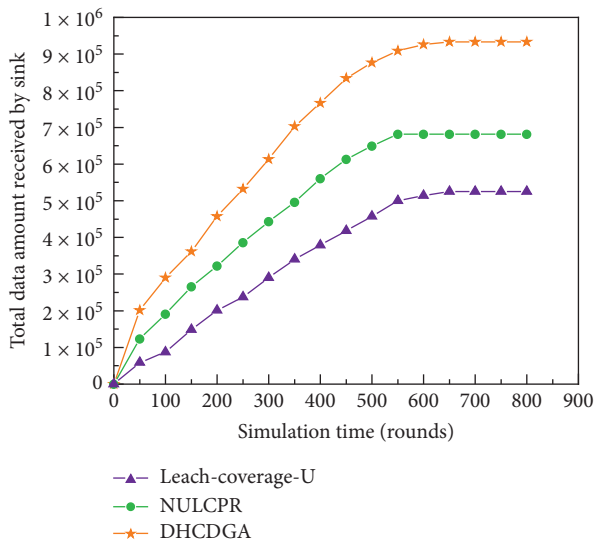


FIGURE 9: Total amount of data received by sink.

clustering algorithm. Therefore, NULCPR receives more data than Leach-coverage-U.

## 7. Conclusions

In this paper, a dynamic hierarchical 3D underwater sensor network clustering data gathering algorithm based on multiple criteria decision making is proposed. Firstly, the entire monitoring network is divided into layers. For selecting a cluster head in each layer, multiple criteria decision making of an intuitionistic fuzzy AHP and hierarchical fuzzy integration is adopted. Furthermore, a sorting algorithm is used to form a clustering topology algorithm to solve the problem that there is the only node in one cluster. Then, an energy-balanced routing algorithm between clusters is proposed according to the residual energy of the node,

the depth, and the number of neighbor nodes. Finally, the simulation results show that DHCDGA can not only effectively balance the energy consumption of the network and prolong the network lifetime but also improve network coverage and data gathering reliability.

Although DHCDGA shows good performances, it does not focus on fault tolerance, security, and other application scenarios in underwater sensor networks. We will study clustering technology in a more comprehensive application scenario in the underwater sensor network.

## Data Availability

The data used to support the findings of this study are included in the article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This research was supported by the National Nature Science Foundation of China under Grant no. 61772101, in part by the Ph.D. Scientific Research Starting Foundation of Liaoning Province, and in part by Innovative Talents in Universities of Liaoning Province under Grant no. LR2019005.

## References

- [1] Z. Liu, B. Hu, B. Huang, L. Lang, H. Guo, and Y. Zhao, "Decision optimization of low-carbon dual channel supply chain of auto parts based on smart city architecture," *Complexity*, vol. 2020, Article ID 2145951, 14 pages, 2020.
- [2] Y. Sun, C. Xu, G. Li et al., "Intelligent human computer interaction based on non redundant EMG signal," *Alexandria Engineering Journal*, vol. 59, no. 3, p. 1149, 2020.
- [3] Z. Chen, D. Chen, Y. Zhang, X. Cheng, M. Zhang, and C. Wu, "Deep learning for autonomous ship-oriented small ship detection," *Safety Science*, vol. 130, Article ID 104812, 2020.
- [4] M. Jouhari, K. Ibrahim, H. Tembine, and J. Ben-Othman, "Underwater wireless sensor networks: a survey on enabling technologies, localization protocols, and internet of underwater things," *IEEE Access*, vol. 7, pp. 96879–96899, 2019.
- [5] N. Goyal, M. Dave, and A. K. Verma, "Data aggregation in underwater wireless sensor network: recent approaches and issues," *Journal of King Saud University—Computer and Information Sciences*, vol. 31, no. 3, pp. 275–286, 2019.
- [6] A. A. Awan, M. A. Khan, A. N. Malik et al., "Quality of service-based node relocation technique for mobile sensor networks," *Wireless Communications and Mobile Computing*, vol. 2019, Article ID 5043187, 13 pages, 2019.
- [7] S. Kumari, P. K. Mishra, and V. Anand, "Fault resilient routing based on moth flame optimization scheme for underwater wireless sensor networks," *Wireless Networks*, vol. 26, no. 2, pp. 1417–1431, 2020.
- [8] N. Javaid, U. Shakeel, A. Ahmad, N. Alrajeh, Z. A. Khan, and N. Guizani, "DRADS: depth and reliability aware delay sensitive cooperative routing for underwater wireless sensor networks," *Wireless Networks*, vol. 25, no. 2, pp. 777–789, 2019.

- [9] W. Zhao, Z. Tang, Y. Yang, L. Wang, and S. Lan, "Cooperative search and rescue with artificial fishes based on fish-swarm algorithm for underwater wireless sensor networks," *The Scientific World Journal*, vol. 2014, Article ID 145306, 10 pages, 2014.
- [10] W. Wei, H. Song, W. Li, P. Shen, and A. Vasilakos, "Gradient-driven parking navigation using a continuous information potential field based on wireless sensor network," *Information Sciences*, vol. 408, no. 2, pp. 100–114, 2017.
- [11] J. Heidemann, M. Stojanovic, and M. Zorzi, "Underwater sensor networks: applications, advances and challenges," *Philosophical Transactions of the Royal Society A*, vol. 370, no. 1958, pp. 158–175, 2012.
- [12] K. Chen, M. Ma, E. Cheng, F. Yuan, and W. Su, "A survey on MAC protocols for underwater wireless sensor networks," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1433–1447, 2014.
- [13] S. Ciment, A. Sanchez, J. V. Capella, N. Meratnia, J. J. Serrano et al., "Underwater acoustic wireless sensor networks: advances and future trends in physical, MAC and routing layers," *Sensors*, vol. 14, no. 1, pp. 795–833, 2014.
- [14] M. Ayaz, A. Abdullah, I. Faye, and Y. Batira, "An efficient dynamic addressing based routing protocol for underwater wireless sensor networks," *Computer Communications*, vol. 35, no. 4, pp. 475–486, 2012.
- [15] M. Moradi, J. Rezazadeh, A. S. Ismail et al., "A reverse localization scheme for underwater acoustic sensor networks," *Sensors*, vol. 12, no. 4, pp. 4352–4380, 2012.
- [16] A. K. Mohapatra, N. Gautam, R. L. Gibson et al., "Combined routing and node replacement in energy-efficient underwater sensor networks for seismic monitoring," *IEEE Journal of Oceanic Engineering*, vol. 38, no. 1, pp. 80–90, 2013.
- [17] N. Z. Zenia, M. Aseeri, M. R. Ahmed, Z. I. Chowdhury, and M. Shamim Kaiser, "Energy-efficiency and reliability in MAC and routing protocols for underwater wireless sensor network: a survey," *Journal of Network and Computer Applications*, vol. 71, pp. 72–85, 2016.
- [18] H. Yan, Z. J. Shi, and J.-H. Cui, "DBR: depth-based routing for underwater sensor networks," in *Proceedings of the 7th International IFIP-TC6 Networking Conference*, pp. 72–86, Singapore, May 2008.
- [19] X. Song, Q. Zhang, W. Sun, and W. Wei, "Energy-efficient data gathering protocol in unequal clustered WSN utilizing fuzzy multiple criteria decision making," *Journal of Intelligent & Fuzzy Systems*, vol. 32, no. 5, pp. 3461–3473, 2017.
- [20] Y. R. Tsai, "Coverage-preserving routing protocols for randomly distributed wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 4, pp. 1240–1245, 2006.
- [21] P. Jiang and X. Wang, "Network layered coverage preserving routing algorithm for underwater sensor networks," *Acta Elect Ronica Sinica*, vol. 44, no. 5, pp. 1240–1246, 2016.
- [22] T. Gurkan, "Clustering-based energy-efficient routing approach for underwater wireless sensor networks," *International Journal of Sensor Networks*, vol. 27, no. 1, pp. 26–36, 2018.
- [23] S. Wang, T. L. N. Nguyen, and Y. Shin, "Energy-efficient clustering algorithm for magnetic induction-based underwater wireless sensor networks," *IEEE Access*, vol. 7, pp. 5975–5983, 2019.
- [24] M. Rao and N. K. Kamila, "Spider monkey optimization based energy efficient clustering in heterogeneous underwater wireless sensor networks," *International Journal of Ad Hoc & Ubiquitous Computing*, vol. 29, no. 1-2, pp. 50–63, 2018.
- [25] M. Faheem, G. Tuna, and V. C. Gungor, "QERP: quality-of-service (QoS) aware evolutionary routing protocol for underwater wireless sensor networks," *IEEE Systems Journal*, vol. 12, no. 3, pp. 2066–2073, 2018.
- [26] K. Bhattacharjya, S. Alam, and D. De, "CUWSN: energy efficient routing protocol selection for cluster based underwater wireless sensor network," *Microsystem Technologies-Micro-and Nanosystems-Information Storage and Processing Systems*, pp. 1–17, 2019.
- [27] V. Krishnaswamy and S. S. Manvi, "Fuzzy and PSO based clustering scheme in underwater acoustic sensor networks using energy and distance parameters," *Wireless Personal Communications*, vol. 108, no. 3, pp. 1529–1546, 2019.
- [28] W. Khan, H. Wang, M. S. Anwar, M. Ayaz, S. Ahmad, and I. Ullah, "A multi-layer cluster based energy efficient routing scheme for UWSNs," *IEEE Access*, vol. 7, pp. 77398–77410, 2019.
- [29] F. Wang, B. Hexiang, L. Deyu, and W. Jianjun, "Energy-efficient clustering algorithm in underwater sensor networks based on fuzzy C means and Moth-flame optimization method," *IEEE Access*, vol. 99, pp. 1–15, 2020.
- [30] E. M. Sozer, M. Stojanovic, and J. G. Proakis, "Underwater acoustic networks," *IEEE Journal of Oceanic Engineering*, vol. 25, no. 1, pp. 72–83, 2000.
- [31] Y. Guo and Y. Liu, "Localization for anchor-free underwater sensor networks," *Computers & Electrical Engineering*, vol. 39, no. 6, pp. 1812–1821, 2013.
- [32] T. Liu, J. Peng, and J. Yang, "Data delivery for heterogeneous delay tolerant mobile sensor networks based on forwarding probability," *Journal of Software*, vol. 24, no. 2, pp. 215–229, 2013.
- [33] C. Zhang, W. Li, and L. Wang, "AHP under the intuitionistic fuzzy environment," in *Proceedings of the 8th International Conference on Fuzzy Systems and Knowledge Discovery*, pp. 583–587, Shanghai, China, July 2011.
- [34] T. Gao, R. C. Jin, J. Y. Song, T. B. Xu, and L. D. Wang, "Energy-efficient cluster head selection scheme based on multiple criteria decision making for wireless sensor networks," *Wireless Personal Communications*, vol. 63, no. 4, pp. 871–894, 2012.

## Research Article

# Hemodynamic-Based Evaluation on Thrombosis Risk of Fusiform Coronary Artery Aneurysms Using Computational Fluid Dynamic Simulation Method

Haoran Wang,<sup>1,2</sup> Hitomi Anzai,<sup>1</sup> Youjun Liu,<sup>3</sup> Aike Qiao ,<sup>3</sup> Jinsheng Xie,<sup>4</sup> and Makoto Ohta <sup>1,5</sup>

<sup>1</sup>Institute of Fluid Science, Tohoku University, 2-1-1, Katahira, Aoba-Ku, Sendai, Miyagi 980-8577, Japan

<sup>2</sup>Graduate School of Biomedical Engineering, Tohoku University, 6-6 Aramaki-Aza-Aoba, Aoba-Ku, Sendai, Miyagi 980-8579, Japan

<sup>3</sup>College of Life Science and Bioengineering, Beijing University of Technology, No. 100 Pingleyuan, Chaoyang District, Beijing 100022, China

<sup>4</sup>Department of Cardiac Surgery, Beijing Anzhen Hospital Capital Medical University, Beijing Institute of Heart Lung and Blood Vessel Diseases, Beijing 100029, China

<sup>5</sup>ELyTMax UMI 3757, CNRS-Université de Lyon, Tohoku University, Sendai, Japan

Correspondence should be addressed to Makoto Ohta; [ohta@biofluid.ifs.tohoku.ac.jp](mailto:ohta@biofluid.ifs.tohoku.ac.jp)

Received 15 May 2020; Revised 13 July 2020; Accepted 20 July 2020; Published 20 October 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Haoran Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Coronary artery aneurysms (CAAs) have been reported to associate with an increased risk for thrombosis. Distinct to the brain aneurysm, which can cause a rupture, CAA's threat is more about its potential to induce thrombosis, leading to myocardial infarction. Case reports suggest that thrombosis risk varied with the different CAA diameters and hemodynamics effects (usually wall shear stress (WSS), oscillatory shear index (OSI), and relative residence time (RRT)) may relate to the thrombosis risk. However, currently, due to the rareness of the disease, there is limited knowledge of the hemodynamics effects of CAA. The aim of the study was to estimate the relationship between hemodynamic effects and different diameters of CAAs. Computational fluid dynamics (CFD) provides a noninvasive means of hemodynamic research. Four three-dimensional models were constructed, representing coronary arteries with a normal diameter (1x) and CAAs with diameters two (2x), three (3x), and five times (5x) that of the normal diameter. A lumped parameter model (LPM) which can capture the feature of coronary blood flow supplied the boundary conditions. WSS in the aneurysm decreased 97.7% apparently from 3.51 Pa (1x) to 0.08 Pa (5x). OSI and RRT in the aneurysm were increased apparently by two orders of magnitude from 0.01 (1x) to 0.30 (5x), and from 0.38 Pa<sup>-1</sup> (1x) to 51.59 Pa<sup>-1</sup> (5x), separately. Changes in the local volume of the CAA resulted in dramatic changes in local hemodynamic parameters. The findings demonstrated that thrombosis risk increased with increasing diameter and was strongly exacerbated at larger diameters of CAA. The 2x model exhibited the lowest thrombosis risk among the models, suggesting the low-damage (medication) treatment may work. High-damage (surgery) treatment may need to be considered when CAA diameter is 3 times or higher. This diameter classification method may be a good example for constructing a more complex hemodynamic-based risk stratification method and could support clinical decision-making in the assessment of CAA.

## 1. Introduction

A coronary artery aneurysm (CAA) is a localized swelling in one of the coronary arteries that supply oxygen to the heart. It is typically defined by a diameter at least 1.5 times that of

the adjacent normal segments of the artery. It is called the giant CAA when the diameter of a CAA exceeds four times the diameter of its adjacent coronary artery [1]. The incidence of CAAs has been reported as 0.3%-5.3% [2]. However, the prevalence of CAA in different literature varies due

to statistical data, [3–6], lack of standardization for diagnosis on angiography [5, 7], and different CAA definitions [8]. CAA is generally associated with the risk of complications such as infarction, ischemia, and thrombosis, with the thrombosis being the most common [9]. Unlike other types of aneurysms, which pose a risk of rupture, the main risk caused by CAA is thrombosis rather than rupture [1].

Current American Heart Association (AHA) guidelines for risk stratification rely on aneurysm size alone as the criterion for initiating systemic anticoagulation; however, studies suggest risk may also depend on hemodynamic parameters such as low wall shear stress (WSS), high oscillatory shear index (OSI), and high relative residence time (RRT), which are not available through image data alone [10–13]. Indeed, CAAs' complex geometries, such as the diameter, may contribute to abnormal hemodynamics and correlate with patient outcome. Considering the other risks, doctors may be confused about the dilemma of whether to adopt the watch-and-wait approach, with the risk of sudden deterioration, or to expose the patient to the risks of treatment if just based on the image data alone. Therefore, it is necessary to evaluate the relationship between CAA diameter and level of thrombosis risk based on the hemodynamic study which could provide support for the doctors.

Patient-specific hemodynamic simulations by using computational fluid dynamics (CFD) can non-invasively supply informative hemodynamic parameters for better thrombotic risk assessment; however, existing studies have been limited and struggle to quantify effects of aneurysm shape on local hemodynamics. With such limitations in mind, existing patient-specific models can be augmented by introducing artificial aneurysms of specified diameter, to achieve a systematic evaluation of the relationship between CAA size and the local hemodynamics. Ultimately, elucidating the relationship between hemodynamics and aneurysmal size characteristics may underlie more powerful risk stratification methods to support clinical decision-making. Usually, these treatments include high-risk surgical treatment and conservative treatment regarding initiation of anticoagulant therapy. Surgery has multiple risks such as the risk of anesthesia, postoperative complications, and the like. Due to the large changes in the cardiovascular structure of the operation, even if the surgical operation has complete risk management, some postoperative complications such as weakened cardiac function and arrhythmia cannot be completely avoided.

CFD has been used widely in the research of vascular diseases in recent years [14–16] and can be used to estimate important hemodynamic parameters including mass flow, WSS, OSI, and RRT. The mass-flow rate can usually be calculated based on speed measurement using non-invasive ultrasonic velocimetry. However, other parameter, WSS, OSI, and RRT, measurements are much more complicated and cannot be accurately measured directly. The indirect calculation method of in vitro experiments, which means the need to build an experimental system and need to consider every potential problem in the system, also has high cost. The advantages of CFD, such as gaining insight into systems that might be difficult to test through

experimentation and lower costs, can overcome the shortcomings of other measurement methods mentioned above. Simulations have played important roles in other studies, such as bypass grafts [17, 18] and abdominal and cerebral aneurysms [19–21], but this technique has rarely been applied to CAAs. One of the important factors is that the characteristics of the coronary artery make its simulation difficult. In the human physiological environment, the resistance of the coronary arteries changes repeatedly; during systole, the resistance of the coronary circulation increases considerably due to the contraction of the heart. In contrast, during diastole, the pressure decreases, which reduces coronary microcirculation resistance. As a consequence, the coronary blood supply increased. In order not to lose the characteristics of the coronary artery, we have adopted a circuit model called lumped parameter model (LPM) to simulate the coronary artery to provide the boundary conditions for the CFD simulation work [22].

The aim of the study was to estimate the relationship between hemodynamic effects and different diameter of CAAs. We quantified hemodynamic parameters of likely clinical relevance for assessing the thrombotic risk including WSS, OSI, and RRT. We created virtual models by artificially constructing models of normal coronary and different diameter of CAAs to compare normal and pathological hemodynamics. This study will allow us to estimate the risk of thrombosis of CAAs from a hemodynamic perspective and provide more comprehensive support for doctors' decisions.

## 2. Materials and Methods

*2.1. Model Development and Simulation Methods.* A patient-specific anatomy model was constructed, based on patient data provided by Anzhen Hospital. All patient information for this analysis was anonymized. This study has been approved by the Medical Ethics Committee of Beijing Anzhen Hospital and Tohoku University. The images used for the 3D model reconstruction were obtained by CT. A personalized real 3D anatomic geometry model was built with manual and threshold segmentation by using Mimics processing software (Materialise NV, Leuven, Belgium). The single fusiform CAAs of different diameters were added to the left anterior descending artery by using the PHANTOM Desktop in Freeform software (3D Systems, USA). We have only changed the diameter of CAA and have not changed the length of CAA along the coronary artery. The link between CAA and normal coronary artery was smoothed. All the changes have been confirmed by the medical doctor. CAA models were constructed with diameters two, three, and five times that of the normal segments of the coronary artery, as well as a model of a normal coronary artery without an aneurysm (Models 2x, 3x, and 5x, respectively; Figure 1), so that we can study hemodynamics on both regular and giant size (Figure 1).

The computational models were generated by 3D model meshing, using a hexahedral mesh controlled by size control using ANSYS Meshing software (Swanson Analysis Systems, Inc., USA). The grids of the regions local to the CAA were refined to allow more precise simulation calculations in these areas for more accurate results.

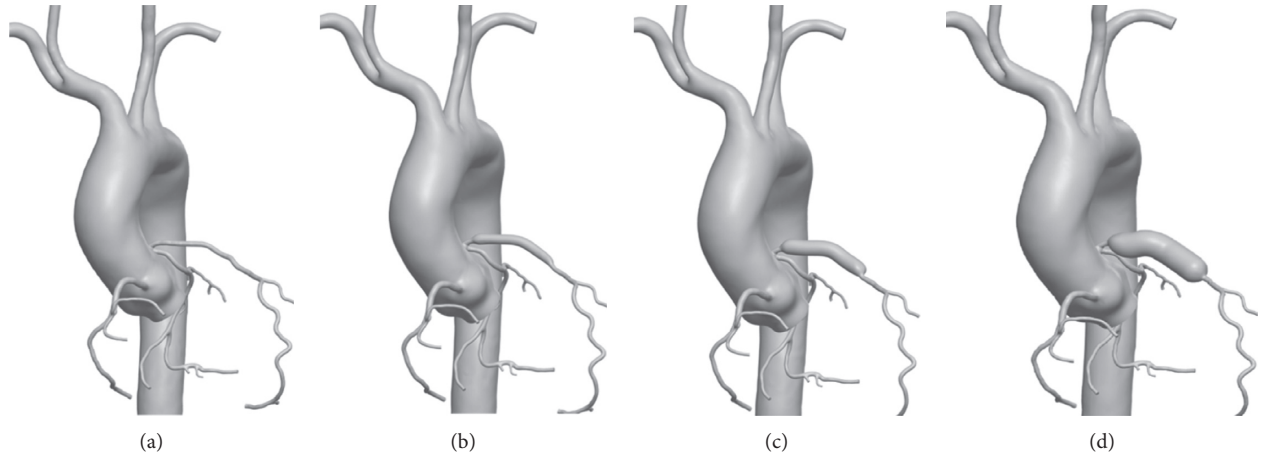


FIGURE 1: Three-dimensional models with a normal coronary artery (Model 0) and with coronary artery aneurysms with diameters two, three, and five times that of the normal diameter (Models 2x, 3x, and 5x, respectively). (a) Model 0. (b) Model 2x. (c) Model 3x. (d) Model 5x.

To confirm the reliability of the simulation results, we performed a grid independence test on the number of elements and nodes [22]. See Figure 2 to check more information of the independence test. It can be found that the simulation results tend to be stable when the number of grids exceeds a certain standard. The total numbers of nodes and elements for the four models differed (Table 1), but they were all of the same order of magnitude number. Blood flow simulations were performed by using the Navier–Stokes solver in the ANSYS-CFX software (Swanson Analysis Systems, Inc.). The simulation calculation assumed that the artery wall was rigid and stationary, the blood flow was Newtonian, and the density and viscosity were  $1050 \text{ kg/m}^3$  and  $0.0035 \text{ Pa s}$ , respectively. Based on the Reynolds number, the fluid is set as laminar flow. The simulations were run for three cardiac cycles. We selected a time step of  $0.0025 \text{ s}$  to satisfy the calculation of stability conditions. The maximum residual of momentum and continuity of the iterative calculation was set to  $0.0001$ .

**2.2. Lumped Parameter Boundary Conditions.** The coronary artery blood flow simulation required careful consideration of the physiological function of the coronary circulation and microcirculation, especially the drastic change in microcirculatory resistance due to heart contractions. This could be simulated appropriately by using an LPM. The LPM construction method for the boundary conditions was based on previous research [22, 23]. The LPM system comprised a heart left ventricle module, coronary artery modules, and arterial modules. These modules were developed based on the concept of the three-element Windkessel RCR (resistor-capacitor-resistor) model [24], a simple circuit model in which blood flow and blood pressure were considered to be analogous to electrical current and voltage, and the resistance and compliance of blood vessels were characterized by resistance and capacitance, respectively. The entire multiscale model is shown in Figure 3. The heart module can be built using some components such as power source and variable capacitor on the basis of the RCR model. The power can simulate the

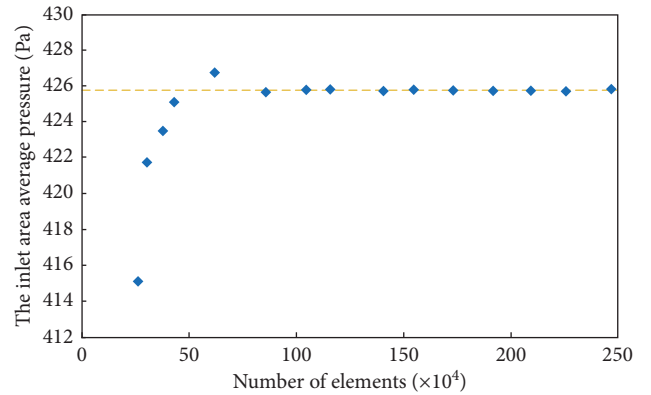


FIGURE 2: The result of grid independence test and the trend line of the results.

TABLE 1: The numbers of nodes and elements of the grid in the four computational models.

|          | Model 0   | Model 2x  | Model 3x  | Model 5x  |
|----------|-----------|-----------|-----------|-----------|
| Nodes    | 1,098,996 | 940,577   | 1,047,183 | 936,749   |
| Elements | 1,437,901 | 1,231,913 | 1,370,433 | 1,230,434 |

blood pressure and the variable capacitor can simulate the volume change of the ventricular during the systole and diastole. The inlet flow via the ascending aorta was provided by the LPM heart module, and the LPM heart model was connected to the LPM coronary artery models through intramyocardial pressure. The components of this module included resistors, capacitors, inductors, and diodes, which were used to simulate coronary resistance, vascular compliance, blood flow inertia, and the heart valves, respectively. The pressure–volume relationship representing the vascular compliance was calculated as follows:

$$C(t) = \frac{V(t) - V_0}{P(t)} = \frac{1}{E(t)}, \quad (1)$$

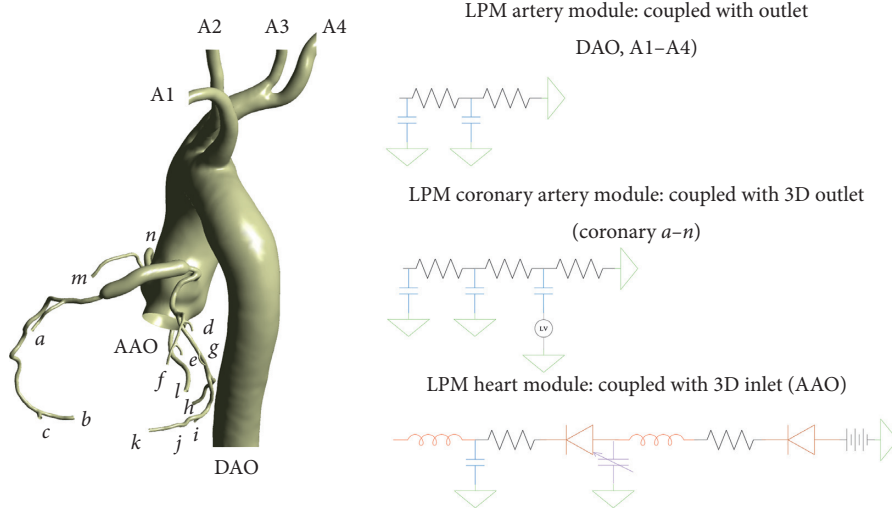


FIGURE 3: Details of the multiscale model used in the study, showing the coupled interface between the lumped parameter model (LPM) and the three-dimensional model. *an*, coronary arteries; A1A4, arteries; AAO, ascending aorta; DAO, descending aorta; RCR, resistor-capacitor-resistor.

where  $V(t)$  and  $P(t)$  are the pressure and volume of the ventricle and  $V_0$  is the reference pressure.  $E(t)$  characterizes the time-varying elastic properties and was calculated as follows:

$$E(t) = E_{\min} + (E_{\max} - E_{\min}) \cdot E_n(t_n). \quad (2)$$

The units of  $E$  used were mL and mmHg. Normalized time-varying elasticity  $E_n(t_n)$  was calculated using a method based on previous research [25] as follows:

$$E_n(t_n) = 1.55 \left[ \frac{1}{1 + (t_n/1.17)^{21.9}} \right] \left[ \frac{(t_n/0.7)^{1.9}}{1 + (t_n/0.7)^{1.9}} \right]. \quad (3)$$

In the present study, we set  $E_{\min} = 0.002458$  and  $E_{\max} = 2.0$ .  $t_n = (t / (0.2 + 0.15t_c))$ . The cardiac cycle  $t_c$  was set to 0.8 s.

Compared with normal arterial blood vessels, the coronary artery module required the addition of a microcirculation module to take account of the special physiology of coronary arteries. This module comprised a resistor ( $R_{v\text{-micro}}$ ) and a capacitor ( $C_{i\text{-micro}}$ ), with the cathode of the capacitor coupled to the internal pressure of the heart module to simulate changes in systolic pressure. The target mean flow of every branch of the LPM coronary arteries was determined by the area of each branch [23, 26]. Based on previous studies [22, 27], we assumed that the total perfusion of the coronary arteries was 4% of the cardiac output and that the total perfusion of the left and right coronary arteries was 60% and 40% of the total perfusion of the coronary arteries, respectively. As shown in Figure 2, the simulation of coronary microcirculation is to add a module on the basis of the common artery to simulate the pressure from the left ventricle and the resistance change caused by the pressure change. The debugging of the entire circuit model parameters is based on previous research [22], in which the pressure adjustment needs to match the patient's blood pressure.

The RCR model can be used to directly simulate arterial blood vessels because these do not undergo drastic changes in resistance. The RCR boundary conditions were set as the outlets of the descending aorta and supra-aortic arteries.

The CFD boundary conditions provided by the LPM for the inlet and outlets of the computational model corresponded to different modules, known as the 0D/3D multiscale coupled method [22, 23]. The important reasons for adopting LPM are the easy-to-deploy circuit model and the characteristics of not requiring a mainframe computer. The easy-to-deploy circuit model means that a complex circuit model composed of many different components can be simulated by programming. All four 3D models shared the same LPM, because, apart from the CAA lesion, the peripheral vascular geometry did not change between the models; this ensured changes in the geometric structure of the CAA were the main cause of the different simulation results. Clinical observation has shown that the coronary structure of the distal end of the coronary artery is not affected by the CAA; this was important evidence supporting the selection of this method.

### 3. Results

**3.1. Mass Flow and Pressure Waveforms.** Figure 4 shows the computed mass flow and pressure over a cardiac cycle for the ascending aorta (the inlet to the model), the descending aorta, an outlet artery, and a coronary artery branch (examples of outlets from the model) for the four 3D models. The results confirmed that, except for the mass flow at the coronary artery boundary, mass flow in the outlets peaked during systole. The coronary mass flow peaked during diastole. These results confirmed that the expected life-like characteristics of the coronary circulation, which blood supply mainly occurred during diastole, were successfully captured in the model [22]. High systolic intra-myocardial pressure resulted in increased distal downstream resistance



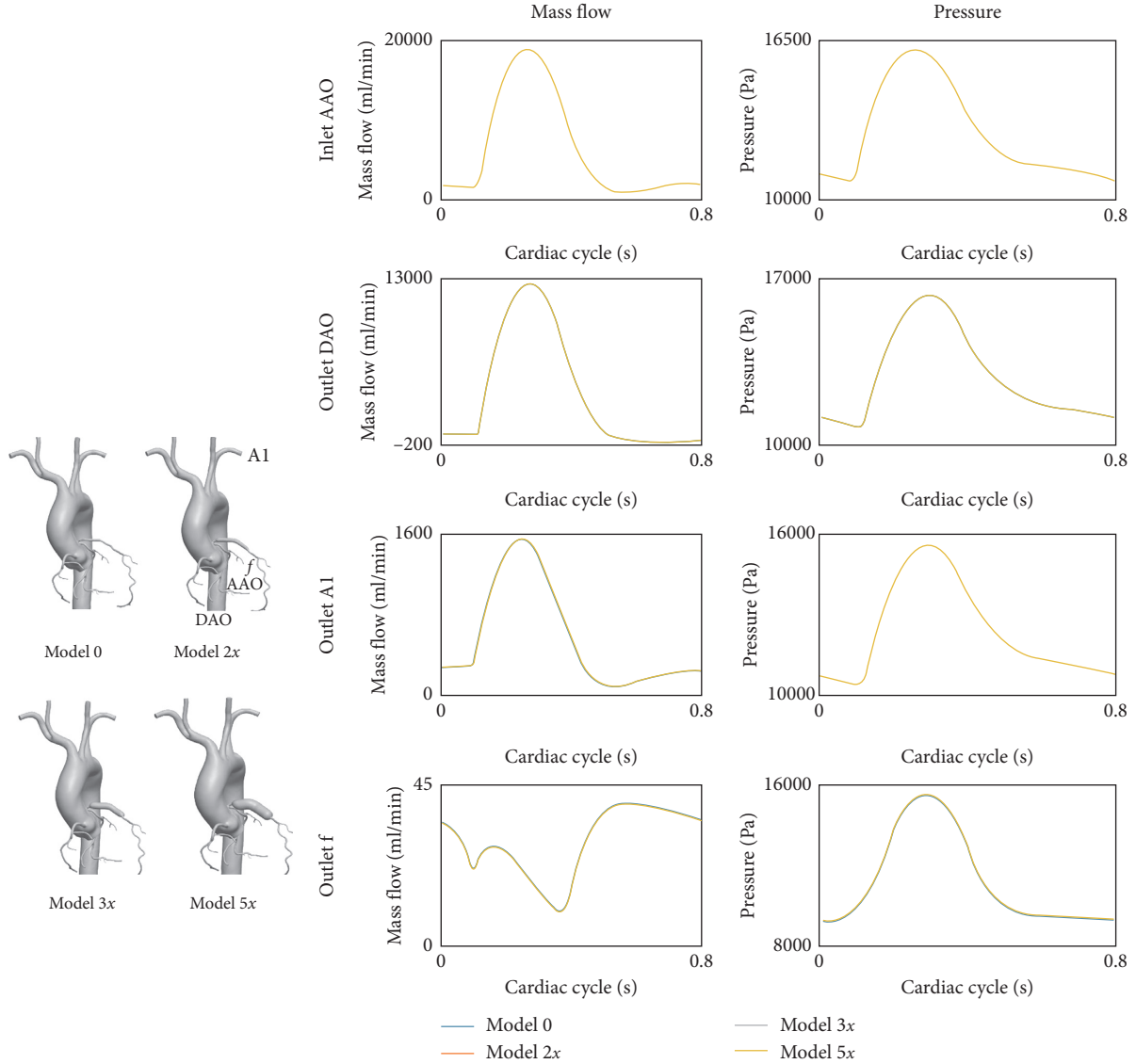


FIGURE 4: Mass flow and pressure curves for the inlet (ascending aorta, AAO) and three outlets (descending aorta (DAO), artery 1 (A1), and coronary artery  $f$ ) over a cardiac cycle.

and decreased blood flow through the coronary artery; the opposite was the case during diastole. Changes in the diameter of the CAA did not have a significant effect on coronary outlet flow.

**3.2. Wall Shear Stress.** The WSS is the force per unit area exerted by the wall on the fluid in a direction along the plane of the local tangent [28], calculated as follows:

$$\tau_w = \frac{4\mu Q}{\pi R^3}, \quad (4)$$

where  $\mu$  represents the blood viscosity,  $Q$  is mass-flow rate, and  $R$  is vessel radius. Previous studies have reported an association between thrombosis formation and low WSS [29]. The area average values of WSS in coronary artery aneurysm over a cardiac cycle in the four models

are listed in Table 2, and the values over a cycle and distribution contours are shown in Figures 5 and 6. The average WSS over a cycle in Model 0 (with no CAA) was 3.51 Pa. The averages for Models 2x, 3x, and 5x were 0.68, 0.26, and 0.08 Pa, respectively, substantially reduced in the CAA due to blood flow stagnation and recirculation, with the reduction increasing in CAAs of greater diameters. The distribution contours in Figure 6 confirm that the CAA results only in local hemodynamic disturbances without wider hemodynamic changes in the arteries. At some distance downstream in the coronary artery, the coronary diameter and WSS both return to normal levels.

**3.3. Oscillatory Shear Index.** OSI is a parameter that quantifies the magnitude of the WSS and directional changes [18], calculated as follows:

TABLE 2: Average values over a cardiac cycle for wall shear stress (WSS), the oscillatory shear index (OSI), and relative residence time (RRT) in the coronary artery aneurysm.

|                                      | Model 0 | Model 2x | Model 3x | Model 5x |
|--------------------------------------|---------|----------|----------|----------|
| Area average WSS (Pa)                | 3.51    | 0.68     | 0.26     | 0.08     |
| Area average OSI                     | 0.01    | 0.03     | 0.13     | 0.30     |
| Area average RRT (Pa <sup>-1</sup> ) | 0.38    | 2.88     | 23.12    | 51.59    |

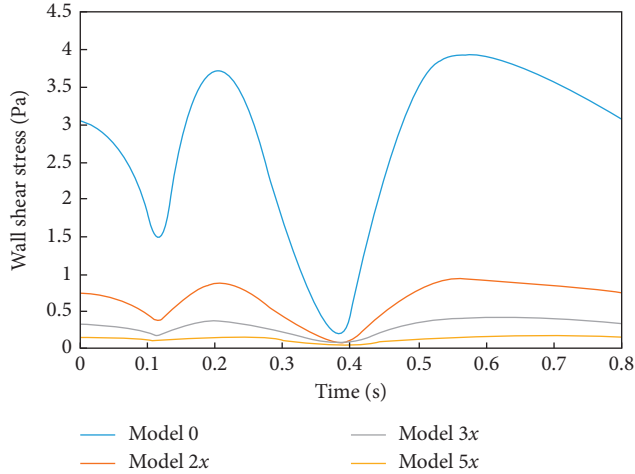


FIGURE 5: Wall shear stress in the coronary artery aneurysm over a cardiac cycle for the four models.

$$\text{OSI} = 0.5 \times \left( 1 - \frac{\left| \int_0^T \tau_\omega dt \right|}{\int_0^T |\tau_\omega| dt} \right), \quad (5)$$

where  $T$  is the cardiac cycle. OSI has values in the range 0.0 to 0.5, with 0 representing undisturbed flow with no change in the direction of the shear stress and 0.5 representing disturbed flow with oscillating shear stress. Previous studies have reported an association between high OSI and the formation of intimal hyperplasia [30]. The average values of OSI in the aneurysm and OSI distribution contours are shown in Table 2 and Figure 7, respectively. The average value of OSI over the cardiac cycle in Model 0 was 0.01. The averages for Models 2x, 3x, and 5x were substantially higher, at 0.03, 0.13, and 0.30, respectively. Previous studies have pointed out that when OSI is greater than 0.2, it will cause endothelial cell damage [31]. As shown in Table 3, the ratio of high OSI area in the aneurysm to the area of aneurysm was 0%, 12.5%, 25.3%, and 66.2%, respectively. Model 0 has OSI close to zero throughout the cardiac cycle, and the flow in Model 0 is unidirectional without circulation. Normal coronary blood flow is unidirectional, usually without flow recirculation, so the OSI of the normal coronary artery (Model 0) was close to zero throughout the entire cardiac cycle. The high OSI values in the CAA models, which reached 30 times higher in model 5x, were due to the increased storage time resulting from the increased size of the cavity, which led to recirculation within the CAA lesion. The effect on OSI was locally confined; there

was little difference between the OSI in the distal coronary arteries in Models 2x, 3x, and 5x and those in Model 0.

**3.4. Relative Residence Time.** The RRT is an indicator that comprehensively characterizes the values and oscillation of WSS [32], as well as the time of local blood flow and vascular interaction [33]. It is calculated as follows:

$$\text{RRT} = \frac{1}{(1 - 2 \cdot \text{OSI}) \cdot \text{TAWSS}}, \quad (6)$$

where

$$\text{TAWSS} = \frac{\text{WSS}}{T}. \quad (7)$$

Here, TAWSS is WSS average over one cardiac cycle, and  $T$  is a cardiac cycle. Figure 8 shows the distribution contours for RRT in the four models, and Table 2 shows the area average values of RRT within the CAA lesion over the cardiac cycle. The average value of RRT in Model 0 was 0.38 Pa<sup>-1</sup>. The averages for Models 2x, 3x, and 5x were substantially higher, at 2.88, 23.12, and 51.59 Pa<sup>-1</sup>, respectively, showing a dramatic increase in RRT with increasing diameter of the CAA. Studies have reported a significantly increased risk of thrombosis when RRT is >10 Pa<sup>-1</sup> [12, 34]. As shown in Table 3, the ratio of high RRT area in the aneurysm to the area of aneurysm was 0%, 0.5%, 27.1%, and 78.6%, respectively. The increased diameter of the CAA makes it easier for blood to remain resident within it, hampering the blood flow to the distal coronary artery. The change in RRT was locally confined to the CAA; there was little difference between the RRT in the distal coronary arteries in Models 2x, 3x, and 5x and that in the distal coronary artery in Model 0.

## 4. Discussion

We have shown that the relationship between hemodynamic effects and different diameters of CAAs can be achieved by using CFD simulation method direct comparison with virtual control case based on the same patient. We found that changes in the diameter of CAA have a strong effect on local hemodynamics while they do not have a significant effect on global hemodynamics. The potential risk of thrombosis increases significantly with increasing CAA diameter indicated through decreasing WSS, increasing OSI and RRT. This research will allow us to estimate the thrombosis risk of CAAs from a hemodynamic perspective and provide more comprehensive support for doctors' decisions.

When making treatment decisions about CAAs, doctors typically depend on their own clinical experience [35]. This may be because of the lack of useful diagnostic tools. The current used tool is the method of imaging which can provide geometric information and almost no other information. The lack of other information cannot provide more help for the doctor's diagnosis. Therefore, it is necessary to find a way to provide more information to help the doctor. It has been shown that WSS over aneurysmal regions is

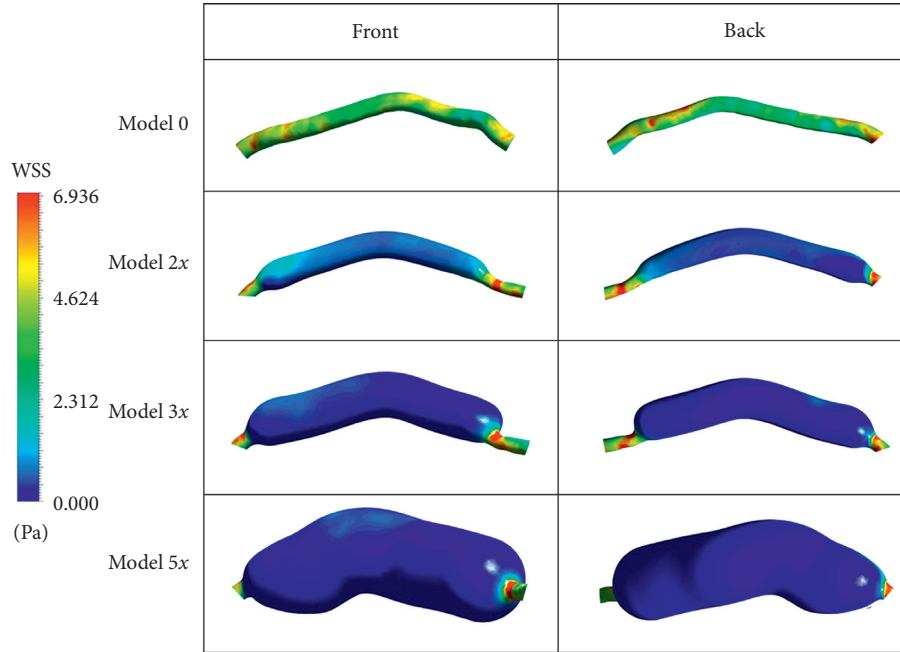


FIGURE 6: Wall shear stress (WSS) distribution in the coronary artery aneurysm at its peak ( $t=0.2$  s) for the four models.

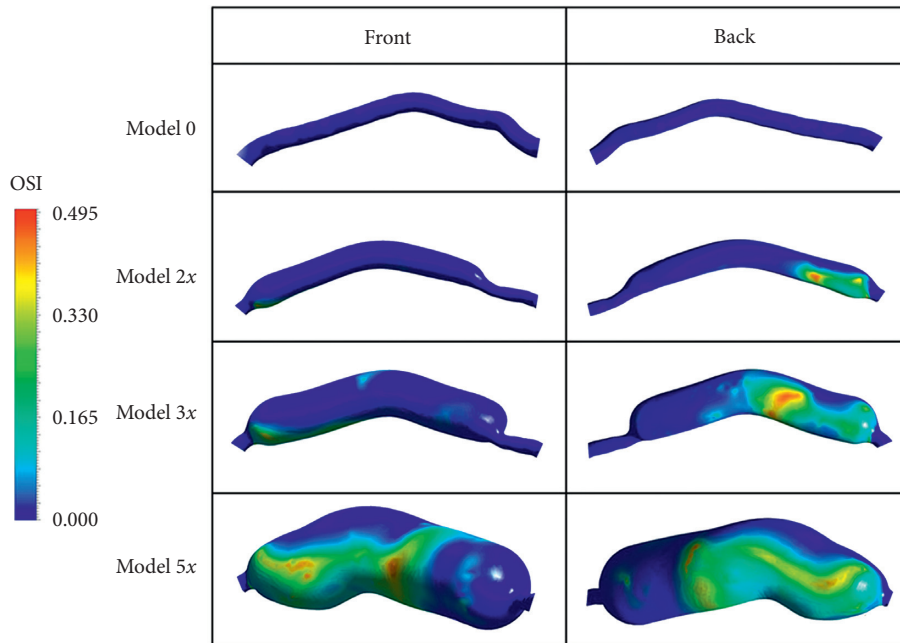


FIGURE 7: Oscillatory shear index (OSI) distribution in the coronary artery aneurysm for the four models.

TABLE 3: The area of high OSI and RRT in the aneurysm and the area of aneurysm.

|  | Model 0       | Model 2x             | Model 3x               | Model 5x                |
|--|---------------|----------------------|------------------------|-------------------------|
| High OSI area/CAA area ( $\text{mm}^2$ ) (ratio) | 0/463.62 (0%) | 95.29/762.38 (12.5%) | 298.53/1179.97 (25.3%) | 1413.89/2135.79 (66.2%) |
| High RRT area/CAA area ( $\text{mm}^2$ ) (ratio) | 0/463.62 (0%) | 38.12/762.38 (0.5%)  | 319.77/1179.97 (27.1%) | 1678.73/2135.79 (78.6%) |

associated with risk of thrombosis, and that is a more predictive factor of thrombosis than aneurysm diameter [36]. We confirmed that hemodynamic research is beneficial for aneurysm diagnosis and treatment [37]. Therefore, in

this study, we quantitatively calculated the hemodynamic parameters of CAA of different diameters in order to estimate the relationship between the risk level of thrombosis and the size of CAA. The results showed that CAAs only

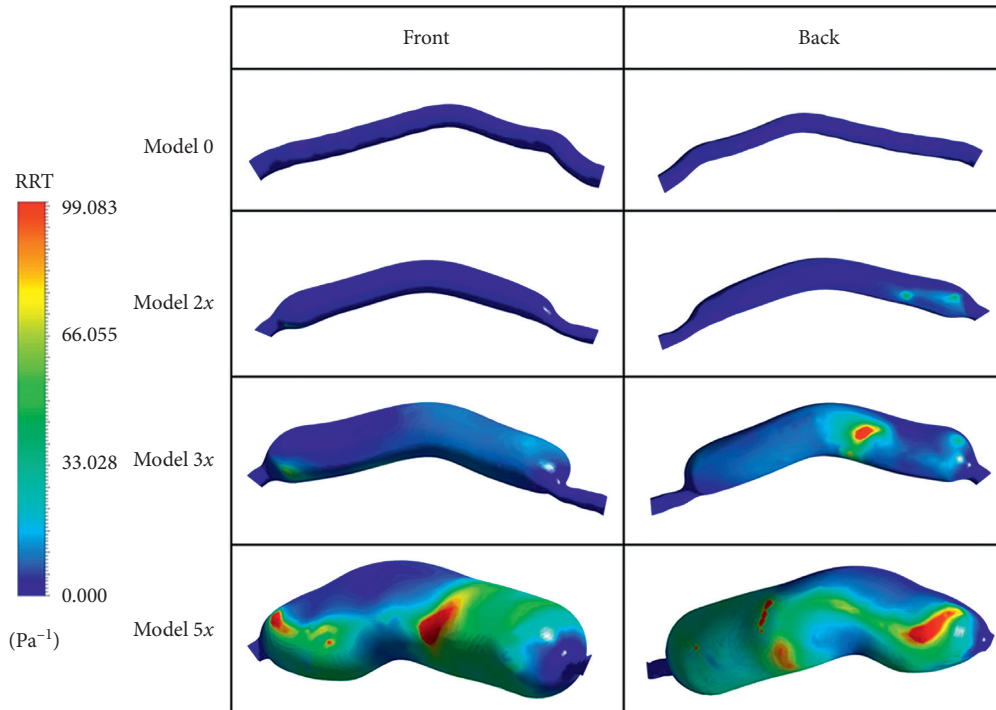


FIGURE 8: Relative residence time (RRT) distribution in the coronary artery aneurysm for the four models.

affect the blood flow and pressure within the region of the lesion itself, with little effect on global blood flow. However, there were substantial changes in local hemodynamic parameters related to thrombosis in the lesion location. Compared with the model without a lesion, all three CAA models showed low WSS, high OSI, and high RRT, with these changes exacerbated as the CAA diameter increased. These changes were the results of the sudden increase in the diameter of the lumen on entering the CAA. These results suggest there may be a high correlation between CAA diameter and the risk of thrombosis. CAA hemodynamic studies may therefore allow the development of a preliminary classification of the risk of thrombosis, helping to improve the effectiveness of anticoagulation therapy and facilitating better personalization of treatment according to a quantitative hemodynamic plan.

Simulation provides a powerful and effective method for estimating important hemodynamic parameters that are difficult to measure. As mentioned above, except mass-flow rate which can be easily detected by non-invasive ultrasound, other parameters, WSS, OSI, and RRT, cannot be directly measured which are usually obtained through in vitro experiments or indirect measurements. Simulation is a non-invasive way, providing a low cost way and avoiding the potential risk of some invasive ways to obtain hemodynamic parameters.

LPM boundary conditions can provide realistic physiologic flow conditions, as what has been confirmed in previous studies [38, 39]. In addition, the coronary microcirculation was introduced into the downstream boundary condition via the LPM. Resistance and pressure changes downstream of the coronary artery could be

accurately captured in this way, helping to ensure the accurate simulation of physiological conditions in the coronary artery. There was no difference between the models in the peripheral vascular structure, so the CFD models shared the same LPM, minimizing the impact of factors other than the diameter of the CAA. Thus, the differences in results between the models were due to the differences in CAA diameters.

As the diameter of the CAA increased, the WSS in the CAA lesion area progressively decreased, with the WSS in Model 5x observed to be >10 times lower than that of the normal coronary artery (Model 0). Parallel plate flow experiments have shown that a large number of platelets accumulate at locations with low WSS (<0.4 Pa) [40], and animal endothelial cell experiments have shown that thrombosis is associated with a decrease in WSS to within the range 0.0770.279 Pa [10]. In the present study, the area average values for WSS in Models 3x and 5x were within this range, suggesting that CAAs with diameters >3 times that of the normal artery are likely to carry a high risk of thrombosis. Although the area average WSS for Model 2x was much lower than that of the model without CAA, it was higher than 0.4 Pa, suggesting that CAAs with diameters double that of the normal artery may not be at high risk of thrombosis. Another important pathophysiological factor in vascular disease is endothelial dysfunction. Low WSS can locally inhibit the production of certain important substances such as NO [41]. An abnormal response to vasodilation agents, including the deactivation of NO by free radicals due to the impairment of acetylcholine or endothelial NO synthase, can lead to abnormal vasoconstriction, which can be associated with endothelial dysfunction [41]. In addition, platelets are prone to aggregation in areas of low WSS, which further

exacerbates the risk of thrombosis. WSS in the CAA lesion of Model 2x was lower than that in the normal model; however, further development is still to be studied just taking WSS as a reference. The risk for thrombosis in CAAs progresses with increasing diameter, so the risk associated with Model 2x would be lower than that for Models 3x and 5x. The development of CAAs needs further study. Although studies have investigated the association between WSS and thrombosis, the current understanding is not yet comprehensive. Therefore, the possibility of WSS and thrombosis in the corresponding *in vitro* experiments should be the focus of continued research.

The OSIs for the CAAs increased rapidly with increasing diameter of the lesion, with the OSI in Model 5x around 30 times greater than that of the normal coronary artery (which was close to zero throughout the CAA). Previous studies have pointed out that when OSI is greater than 0.2, it will cause endothelial cell damage [31]. The high OSI area (OSI > 0.2) ratio increased from 0.1% to 66.2% between Model 0 and Model 5x. *In vivo* experimental studies of porcine aortas have reported that disturbed flow (i.e., an increase in OSI) promoted the expression of proinflammatory cytokines in endothelial cells [11]. In addition, an *in vitro* study of bovine endothelial cells reported that the expression of monocyte chemoattractant protein-1 could be induced by applying a low-amplitude oscillating shear stress of 0.3 Pa at 1 Hz (equivalent to an OSI of 0.5), and this hemodynamic environment enhanced the binding of endothelial cells to monocytes [42]. These findings suggest that the elevated OSI values observed in the CAA models may be associated with conditions that can trigger thrombosis.

The RRT values for the CAAs increased dramatically with increasing diameters of the lesions, with the RRT for Model 5x > 100 times greater than that of the normal coronary artery. The high RRT area ratio increased from 0.1% to 78.6% between Model 0 and Model 5x. These findings indicate that the blood flow in the CAA recirculates when the diameter of the lumen increases, resulting in prolonged contact time between the CAA wall and the blood. Studies have reported a significantly increased risk of thrombosis when RRT is  $>10 \text{ Pa}^{-1}$  [12, 34]. The RRT values for Models 3x and 5x in the present study were both  $>10 \text{ Pa}^{-1}$ , suggesting a high risk of thrombosis when the diameter of a CAA exceeds three times that of the normal coronary artery. Conversely, the RRT for Model 2x was much lower than  $10 \text{ Pa}^{-1}$ , suggesting that the risk of thrombosis may be low when the diameter of the CAA is only double that of the normal artery. Although there have been various studies of the relationship between RRT and aneurysms [43], the current state of knowledge remains insufficient for patient classification; for example, there are no appropriate ranges of RRT values for distinguishing the risk level of thrombosis which needs further study.

In summary, the comparison of our modeling results with the findings of previous studies suggested that the low WSS, high OSI, and high RRT observed in the CAAs would significantly increase the risk of thrombosis. However, the risk varied considerably between the different diameters of CAA. The WSS, OSI, and RRT values for Models 3x and 5x all suggested a high risk of thrombosis, based on the previous study results. Thus, for CAAs with diameters >3 times that of

the normal coronary artery, the risk of uncertainty with conservative treatment may be higher than the risk associated with surgery, suggesting that the surgical operation should be considered. Conversely, although the WSS, OSI, and RRT values for Model 2x were unfavorable compared with those of the model of the normal coronary artery, implying an increased risk of thrombosis, the magnitude of these differences was insufficient to support a prediction of the future development of the CAA. For CAAs with a diameter up to two times that of the normal coronary artery, conservative drug treatment may be the better option, avoiding the risks associated with surgery.

The limitations of this study are as follows. First, this study has only one patient's data, and the conclusion may be accidental. We need to use more data for research. Secondly, an important assumption used in the CFD calculation is the rigid wall, where, in fact, elastic walls should be used. In particular, the coronary artery will deform greatly as the heart contracts which can be solved by two-way FSI in the future.

## 5. Conclusion

The evaluation of the relationship between hemodynamic effects and different diameters of CAAs was performed in this research through the CFD simulation method. The relationship between thrombosis risk and the size of the CAA diameter could be preliminarily evaluated from a hemodynamic perspective.

We found that as the diameter of the CAA increased, the risk of thrombosis increased substantially. The results suggested that when the diameter of a CAA is small (less than double that of the normal coronary artery), the risk of thrombosis may be relatively low, suggesting that medication may be a sufficient approach for the patient, avoiding the uncertain risks associated with surgery. If the diameter of CAA is large (>3 times that of the normal coronary artery), there may be a high risk of thrombosis, and surgical treatment could be considered. A similar approach could be taken in future studies to systematically assess the risk associated with different shapes, sizes, and locations of aneurysms. Such approaches may form the basis for more sophisticated hemodynamic-based risk stratification methods supporting clinical decision-making in assessment of CAA.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

This research was supported by the National Natural Science Foundation of China (nos. 11832003 and 11772016) and JSPS KAKENHI under Grant no. JP18K18355. This research was partially supported by the Creation of a Development

Platform for Implantable/Wearable Medical Devices by a Novel Physiological Data Integration System of the Program on Open Innovation Platform with Enterprises, Research Institute and Academia (OPERA) from the Japan Science and Technology Agency (JST). This research was also supported by Collaborative Research Project 2020, Institute of Fluid Science, Tohoku University (J20R001) with the Grant Number JP18K18355.

## References

- [1] H. Kato, T. Sugimura, T. Akagi et al., “Long-term consequences of kawasaki disease,” *Circulation*, vol. 94, no. 6, pp. 1379–1385, 1996.
- [2] S. A. Sherif, O. O. Tok, Ö. Taşköylü, O. Goktekin, and I. D. Kilic, “Coronary, artery aneurysms: a review of the epidemiology, pathophysiology, diagnosis, and treatment,” *Frontiers in Cardiovascular Medicine*, vol. 4, 2017.
- [3] E. Beckmann, S. Rustum, S. Marquardt et al., “Surgical treatment of coronary artery aneurysms,” *Journal of Cardiac Surgery*, vol. 32, no. 11, pp. 674–679, 2017.
- [4] M. Syed, M. Lesch, and A. D. Syed, “Coronary artery aneurysm: a review,” *Progress in Cardiovascular Diseases*, vol. 40, no. 1, pp. 77–84, 1997.
- [5] G. G. Hartnell, B. M. Parnell, and R. B. Pridie, “Coronary artery ectasia. Its prevalence and clinical significance in 4993 patients,” *British Heart Journal*, vol. 54, no. 4, 1985.
- [6] A. Satran, B. A. Bart, C. R. Henry et al., “Increased prevalence of coronary artery aneurysms among cocaine users,” *Circulation*, vol. 111, no. 19, pp. 2424–2429, 2005.
- [7] O. Topaz, M. S. Rutherford, S. M. Bojack et al., “Giant aneurysms of coronary arteries and saphenous vein grafts: angiographic findings and histopathological correlates,” *Cardiovascular Pathology*, vol. 14, no. 6, pp. 298–302, 2005.
- [8] A. P. Antoniadis, Y. S. Chatzizisis, and G. D. Giannoglou, “Pathogenetic mechanisms of coronary ectasia,” *International Journal of Cardiology*, vol. 130, no. 3, pp. 335–343, 2008.
- [9] J. B. Gordon, A. M. Kahn, and J. C. Burns, “When children with kawasaki disease grow up,” *Journal of the American College of Cardiology*, vol. 54, no. 21, pp. 1911–1920, 2009.
- [10] P. Sriramarao, L. Languino, and D. Altieri, “Fibrinogen mediates leukocyte-endothelium bridging in vivo at low shear forces,” *Blood*, vol. 88, no. 9, pp. 3416–3423, 1996.
- [11] A. G. Passerini, D. C. Polacek, C. Shi et al., “Coexisting proinflammatory and antioxidative endothelial transcription profiles in a disturbed flow region of the adult porcine aorta,” *Proceedings of the National Academy of Sciences*, vol. 101, no. 89, pp. 2482–2487.
- [12] U. Morbiducci, D. Gallo, R. Ponzini et al., “Quantitative analysis of bulk flow in image-based hemodynamic models of the carotid bifurcation: the influence of outflow conditions as test case,” *Annals of Biomedical Engineering*, vol. 38, no. 12, pp. 3688–3705, 2010.
- [13] T. Ohkubo, R. Fukazawa, E. Ikegami, and S. Ogawa, “Reduced shear stress and disturbed flow may lead to coronary aneurysm and thrombus formations,” *Pediatrics International*, vol. 49, no. 1, pp. 1–7, 2007.
- [14] H. Anzai, T. Watanabe, X. Han et al., “Endothelial cell distributions and migration under conditions of flow shear stress around a stent wire,” *Technology and Health Care*, vol. 28, no. 4, pp. 345–354, 2020.
- [15] M. Zhang, Y. Li, X. Zhao et al., “Haemodynamic effects of stent diameter and compaction ratio on flow-diversion treatment of intracranial aneurysms: a numerical study of a successful and an unsuccessful case,” *Journal of Biomechanics*, vol. 58, pp. 179–186, 2017.
- [16] H. Anzai, B. Chopard, and M. Ohta, “Combinational optimization of strut placement for intracranial stent using a realistic aneurysm,” *Journal of Flow Control, Measurement & Visualization*, vol. 2, no. 2, pp. 67–77, 2014.
- [17] J. P. Ku, M. T. Draney, F. R. Arko et al., “In vivo validation of numerical prediction of blood flow in arterial bypass grafts,” *Annals of Biomedical Engineering*, vol. 30, no. 6, pp. 743–752, 2002.
- [18] A. R. Soler, F. Kabinejadian, M. A. Slevin, P. J. Bartolo, and A. Keshmiri, “Optimisation of a novel spiral-inducing bypass graft using computational fluid dynamics,” *Scientific Reports*, vol. 7, no. 1, p. 1865, 2017.
- [19] H. Zhang, L. Li, C. Cheng, and X. Sun, “Clinical value of homodynamic numerical simulation applied in the treatment of cerebral aneurysm,” *Experimental and Therapeutic Medicine*, vol. 4, no. 6, pp. 5401–5494, 2017.
- [20] A. S. Les, S. C. Shadden, C. A. Figueroa et al., “Quantification of hemodynamics in abdominal aortic aneurysms during rest and exercise using magnetic resonance imaging and computational fluid dynamics,” *Annals of Biomedical Engineering*, vol. 38, no. 4, pp. 1288–1313, 2010.
- [21] M. Piccinelli, A. Veneziani, D. A. Steinman, A. Remuzzi, and L. Antiga, “A framework for geometric analysis of vascular structures: application to cerebral aneurysms,” *IEEE Transactions on Medical Imaging*, vol. 28, no. 8, pp. 1141–1155, 2009.
- [22] H. J. Kim, I. E. Vignon-Clementel, J. S. Coogan, C. A. Figueroa, K. E. Jansen, and C. A. Taylor, “Patient-specific modeling of blood flow and pressure in human coronary arteries,” *Annals of Biomedical Engineering*, vol. 38, no. 10, pp. 3195–3209, 2010.
- [23] C. A. Taylor, T. A. Fonte, and J. K. Min, “Computational fluid dynamics applied to cardiac computed tomography for noninvasive quantification of fractional flow reserve,” *Journal of the American College of Cardiology*, vol. 61, no. 22, pp. 2233–2241, 2013.
- [24] I. E. V. Clementel, C. A. Figueroa, K. E. Jansen, and C. A. Taylor, “Outflow boundary conditions for 3D simulations of non-periodic blood flow and pressure fields in deformable arteries,” *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 13, no. 5, pp. 625–640, 2010.
- [25] N. Stergiopoulos, J. J. Meister, and N. Westerhof, “Determinants of stroke volume and systolic and diastolic aortic pressure,” *The American Journal of Physiology*, vol. 270, no. 6, pp. 2050–2059, 1996.
- [26] M. Zamir, P. Sinclair, and T. H. Wonnacott, “Relation between diameter and flow in major branches of the arch of the aorta,” *Journal of Biomechanics*, vol. 25, no. 11, pp. 1303–1310, 1992.
- [27] K. Johnson, P. Sharma, and J. Oshinski, “Coronary artery flow measurement using navigator echo gated phase contrast magnetic resonance velocity mapping at 3.0 tesla,” *Journal of Biomechanics*, vol. 41, no. 3, pp. 595–602, 2008.
- [28] F. P. Salvuccia, C. A. Perazzob, S. Sallesa, J. G. Barrac, and R. L. Armentano, “Influence of inlet conditions in wall shear stress distributions of left coronary arteries in patient-specific simulations,” *Mecánica Computacional*, vol. XXIX, pp. 5953–5960, 2010.
- [29] B. C. Lopez, B. Calvo, and P. Estefanía, *Patient-Specific Computational Modeling*, Springer, Berlin, Germany, 2012th edition, 2012.

- [30] D. N. Ghista and F. Kabinejadian, "Coronary artery bypass grafting hemodynamics and anastomosis design: a biomedical engineering review," *BioMedical Engineering Online*, vol. 12, no. 129, 2013.
- [31] F. P. Glor, F. P. Glor, B. Ariff et al., "Image-based carotid flow reconstruction: a comparison between MRI and ultrasound," *Physiological Measurement*, vol. 25, no. 6, pp. 1495–1509, 2004.
- [32] S.-W. Lee, L. Antiga, and D. A. Steinman, "Correlations among indicators of disturbed flow at the normal carotid bifurcation," *Journal of Biomechanical Engineering*, vol. 131, no. 6, Article ID 061013, 2009.
- [33] H. A. Himburg, "Spatial comparison between wall shear stress measures and porcine arterial endothelial permeability," *AJP: Heart and Circulatory Physiology*, vol. 286, no. 5, pp. 1916–1922, 2004.
- [34] T. Suess, J. Anderson, L. Danielson et al., "Examination of near-wall hemodynamic parameters in the renal bridging stent of various stent graft configurations for repairing visceral branched aortic aneurysms," *Journal of Vascular Surgery*, vol. 64, no. 3, pp. 788–796, 2016.
- [35] H. Zhang, X. Yu, F. Ji, and F. Sun, "Coronary artery aneurysm formation after drug-coated," *Medicine*, vol. 97, no. 45, 2018.
- [36] N. G. Gutierrez, M. Mathew, B. W. McCrindle et al., "Hemodynamic variables in aneurysms are associated with thrombotic risk in children with Kawasaki disease," *International Journal of Cardiology*, vol. 281, pp. 15–21, 2019.
- [37] C. Karmonik, G. Chintalapani, T. Redel et al., "Hemodynamics at the ostium of cerebral aneurysms with relation to post-treatment changes by a virtual flow diverter: a computational fluid dynamics study," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 1895–1898, Osaka, Japan, July 2013.
- [38] J. T. Marcus, H. G. Smeenk, J. P. A. Kuijter, R. J. Van Der Geest, R. M. Heethaar, and A. C. Van Rossum, "Flow profiles in the left anterior descending and the right coronary artery assessed by MR velocity quantification: effects of through-plane and in-plane motion of the heart," *Journal of Computer Assisted Tomography*, vol. 23, no. 4, pp. 567–576, 1999.
- [39] Y. Shi, P. Lawford, and R. Hose, "Review of zero-D and 1-D models of blood flow in the cardiovascular system," *Bio-Medical Engineering Online*, vol. 10, no. 1, p. 33, 2011.
- [40] M. Kroll, J. Hellums, L. McIntire, A. Schafer, and J. Moake, "Platelets and shear stress," *Blood*, vol. 88, no. 5, pp. 1525–1541, 1996.
- [41] J.-J. Chiu and S. Chien, "Effects of disturbed flow on vascular endothelium: pathophysiological basis and clinical perspectives," *Physiological Reviews*, vol. 91, no. 1, pp. 327–387, 2011.
- [42] J. Hwang, M. H. Ing, A. Salazar et al., "Pulsatile versus oscillatory shear stress regulates NADPH oxidase subunit expression," *Circulation Research*, vol. 93, no. 12, pp. 1225–1232, 2003.
- [43] S.-i. Sugiyama, K. Niizuma, T. Nakayama et al., "Relative residence time prolongation in intracranial aneurysms," *Neurosurgery*, vol. 73, no. 5, pp. 767–776, 2013.

## Research Article

# Optimization and Simulation for E-Commerce Supply Chain in the Internet of Things Environment

Liwei Zhu 

*College of Commerce, Guangdong Polytechnic of Science and Technology, Zhuhai 519090, China*

Correspondence should be addressed to Liwei Zhu; 11630019@zju.edu.cn

Received 12 August 2020; Revised 5 October 2020; Accepted 8 October 2020; Published 19 October 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Liwei Zhu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The supply chain optimization of e-commerce is the optimization of the integrated linear management model for e-commerce enterprises based on manufacturers, suppliers, warehouses, distribution centers, and channels. The application of Internet of Things (IoT) technology to the optimization of e-commerce supply chain can help e-commerce enterprises to monitor each product in real time and manage logistics system reasonably; thus, the circulation process of e-commerce product in the supply chain can be monitored and shared, and the product information at each stage of the supply chain can be analyzed and predicted, helping e-commerce companies to estimate the future trends or accident probability. On the analysis of summarizing and analyzing previous works, this article expounded the research status and significance of e-commerce supply chain optimization, elaborated the development background, current status, and future challenges of the IoT technology, introduced the methods and principles of the node distribution model and constraint parameter analysis of logistics supply chain, constructed the IoT-based e-commerce supply chain linear and nonlinear model including competition and risk-control modules, performed the design of IoT-based e-commerce supply chain optimization, and discussed the solutions to procurement robustness optimization and management process optimization. The final simulation analysis showed that the proposed optimization model can effectively predict e-commerce product information in each stage of the supply chain; the IoT-based e-commerce supply chain optimization can enable each node company in the supply chain to improve the quality of information transmission and reduce the risk of information asymmetry in the process of e-commerce exchanges and can quickly and easily respond to the sudden changes of market and environment, thereby promoting them to cooperate with each other for establishing a supply chain network structure with closer business relationships, smoother logistics transportation, more reasonable benefit sharing, and stronger market competitiveness.

## 1. Introduction

Internet of Things (IoT) is such an information-based and intelligent network that utilizes the communication technologies such as Internet to connect sensors, people, and things in a new way to form a connection. Through direct communication between things, the time taken for the entire transfer process is shortened, and the operation efficiency of the entire supply chain is improved so as to achieve the real-time tracking, monitoring, and management in a real sense [1]. The IoT is composed of three parts of sensor network, transmission network, and application network, including four key technologies of radio frequency identification, sensors, cloud computing, and network communication.

The rapid developments and industry applications of the IoT have brought huge opportunities to deepen the development of e-commerce [2]. The application of related technologies and methods in the field of e-commerce will help promote the rapid and healthy development of e-commerce. For example, in the aspect of product management, the use of IoT technology to establish a real-time tracking system for products can enable enterprises to monitor the status of products at any time and effectively manage the quality of the products [3]. The IoT provides an effective information technology support platform for the integration of information and resources in the e-commerce supply chain, tracking, and monitoring of the entire product life cycle. However, consumers still feel insecure about e-commerce



psychologically due to various shortcomings in e-commerce logistics, so it is necessary to study the optimization of the IoT-based e-commerce supply chain [4].

The application of IoT to the e-commerce supply chain optimization can help the e-commerce companies to monitor each product in real time and optimize the management of the logistics system, which can not only supervise and circulate products in the supply chain but also share related information [5]. The e-commerce supply chain information integration and sharing platform can achieve unified standards for internal and external information, interoperable, seamless docking, integration, analysis, and sharing so that each node company in the supply chain can perceive information in a short time. It can analyze and forecast the information of products at various stages of the supply chain, help e-commerce companies to estimate future trends or the probability of accidents, so as to take timely remedial measures or early warnings, greatly improve the ability of e-commerce companies to respond to the market, and improve business management capabilities traditional logistics is mostly confined within the enterprise [6]. With information flow as the control object and the main enterprise as the core, the entire logistics system is controlled from the supply source to the demand source [7]. The e-commerce supply chain is an optimization of integrated linear management models based on manufacturers, suppliers, warehouses, distribution centers, and channels, which includes the optimization of five main related links of production, distribution, and retailing in the context of increasingly sophisticated transportation and warehousing, global division of labor, and corporate transnational and cross-regional operations [8].

On the analysis of summarizing and analyzing previous works, this article expounded the research status and significance of e-commerce supply chain optimization, elaborated the development background, current status, and future challenges of the IoT technology, introduced the methods and principles of the node distribution model and constraint parameter analysis of logistics supply chain, constructed an IoT-based e-commerce supply chain linear and nonlinear model including competition and risk-control modules, performed the design of IoT-based e-commerce supply chain optimization, and discussed the solutions to procurement robustness optimization and management process optimization. The study results of this paper provide a certain reference for further research studies on the e-commerce supply chain optimization in the IoT environment. The detailed chapter arrangement is as follows. Section 2 introduces the methods and principles of the node distribution model and constraint parameter analysis of logistics supply chain; Section 3 establishes an IoT-based e-commerce supply chain model including competition and risk-control modules; Section 4 performs the optimization design of IoT-based e-commerce supply chain and discusses the optimization solutions to procurement robustness and management process; Section 5 performs a simulation experiment and its analysis; Section 6 provides conclusion.

## 2. Methods and Principles

*2.1. Node Distribution Model of Logistics Supply Chain.* The e-commerce supply chain is characterized by the complexity and dynamics of the supply chain itself, and the data and models used in supply chain optimization should also meet the dynamic characteristics [9]. Considering the disturbing effect of the dynamic environment on the supply chain node distribution, the optimized node distribution function is constructed as follows:

$$\frac{\partial \varphi(x)}{\partial x} = \frac{1}{k} \cdot \frac{\partial^2}{\partial x^2} [M(x) \cdot \varphi(x)] + \frac{1}{r} \cdot \frac{\partial}{\partial x} [M(x) \cdot \varphi(x)], \quad (1)$$

where  $M(x)$  is the average trend of measuring the optimization level of supply chain  $\varphi(t)$ ;  $k$  is the coefficient of supply chain risk-control ability;  $r$  is the coefficient of diffusion and convergence of supply chain optimization.

It is assumed that there is no shortage of e-commerce product supplies in the upstream and meets the above-mentioned optimization node distribution function, and then the maximum profit function that the upstream supplier may achieve is as follows:

$$M_x(p_x, q_x) = (p_x - k) \cdot q_x^2 - k \left( \frac{1}{\alpha} \cdot q_x - k \right) p_x^2, \quad (2)$$

where  $p_x$  is the cost bonus coefficient in the IoT environment;  $q_x$  is the output matching coefficient in the IoT environment;  $\alpha$  is the profit loss caused by information asymmetry in the entire e-commerce supply chain.

When the distribution of nodes such as manufacturers and retailers in the e-commerce supply chain shares real data, both parties can get a higher net income increase and at the same time bear the risk of data being leaked by each other, so the profit function of the two can be expressed as follows:

$$M_x[p_x, \varphi(x)] = w \cdot Q \cdot \varphi(x)^2 - c[\varphi(x) + P]p_x^2, \quad (3)$$

$$M_x[p(x), \varphi(x)] = \rho_R \cdot Q \cdot \varphi(x)^2 - E_M[\varphi(x) + P]p(x)^2, \quad (4)$$

where  $w$  is the order price;  $c$  is the unit cost,  $Q$  is the market demand,  $P$  is the market price of the product,  $E_M$  is the estimated value of the single-cycle shared data, and  $\rho_R$  is the probability of leaking shared data.

When the distribution of the nodes of the logistics supply chain does not affect each other, the standard deviation of the aggregate industry demand is less than the sum of the standard deviations ( $\sigma$ ) of the individual regional requirements, indicating that the aggregate effect can reduce the fluctuation of demand to achieve the purpose of reducing safety stocks. At this time, the aggregate demand is normally distributed:

$$\sigma = \sqrt{\varphi(x)^2 \cdot s_i^2 + \frac{1}{\rho} \cdot S_i \cdot \text{cov}(i, j) - R[\varphi(x) + P]p_x^2}, \quad (5)$$

where  $S_i$  is the company's expected value of the market demand in each period,  $\text{cov}(i, j)$  is the company's covariance of the market demand in each period, and  $R$  is the average level of the periodic service. Such enterprises do not participate in actual production and require less investment in infrastructure but have high requirements on management and planning.

**2.2. Constraint Parameter Analysis of Logistics Supply Chain.** The degree of information processing in the e-commerce supply chain can be obtained by using the ratio of the number of times the information is automatically processed in each circulation link of the supply chain to the total number of information processing times and then multiplying them by the importance of each supply chain link to obtain the following [10]:

$$E_i = \sum_{j=1}^n \left[ \frac{1}{\rho} \cdot \frac{X_i}{Y_i} + r_{ij} \cdot E_M[\varphi(x) + P]p(x)^2 \right]^2, \quad (6)$$

where  $X_i$  is the number of times that information is automatically processed on the  $i$ -th supply chain node;  $Y_i$  is the total number of times that information is processed on the  $i$ -th supply chain node;  $r_{ij}$  is a weighted value that is the importance of the information automation of the  $i$ -th node company to the entire supply chain.

It is supposed that there are  $n$  optimization processes in an e-commerce supply chain. Let  $X = (x_1, x_2, \dots, x_n)$  ( $i = 1, 2, \dots, n$ ), and each process is represented by  $m$  indicator sets. The optimization process  $x_i$  is measured by attributes, and its attribute value  $a_{ij}$  is obtained from the actual situation so as to form a decision matrix  $A = (a_{ij})_{n \times m}$ :

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}, \quad (7)$$

where  $n$  is the number of rows in the optimization matrix,  $i$  is the number of optimizations for a particular process, and  $k$  is a criterion. If  $r_{ik}$  represents the optimization level of process  $i$  under criterion  $k$ , the ordinal value of the optimization number  $i$  can be given by

$$O_i = \sum_{k=1}^n \sum_{i=1}^n (n - r_{ik}). \quad (8)$$

It assumed that the revenue function of the upstream supplier is  $W_i(s_i, q_i)$ , where  $s_i$  is the sales price of the supplier and  $q_i$  is the sales volume of the supplier;  $Q_i$  is a single item. The cost of selling products and the indestructible constraint parameter function of the upstream supplier is given by

$$W_i(s_i, q_i) = (s_i - Q_i)q_i. \quad (9)$$

The core company's demand in each period follows a normal distribution with an average value of  $m$ ; each supply interval period is  $L$ , so the demand in the  $L$  period also

follows a normal distribution with an average value of  $Lm$  of target inventory holding  $N$  is as follows:

$$N = (LQ_i)^{1/2} \left( \frac{P-C}{P-H} \right) + Lm - W_i(s_i, q_i), \quad (10)$$

where  $P$  is the price provided by the supplier management enterprise to the core enterprise,  $C$  is the unit production cost, and  $H$  is the unit inventory holding cost. When the supplier management company cannot meet the inventory requirements due to some force majeure, the company can only go to resource pool for assistance.

### 3. IoT-Based E-Commerce Supply Chain Model

**3.1. Competition Model of E-Commerce Supply Chain.** In the traditional model, there are information barriers between logistics providers and they rely on information barriers to obtain competitiveness due to the immature information sharing mechanism and technology. But the IoT brings technical support for information sharing. In this context, the reasonable requirements for the disclosure of non-commercial confidential basic information such as distribution customers, channels, and partners of all logistics suppliers have been realized, and they have not obeyed the market because of the market's need for higher quality and lower cost logistics services. In the distribution and retail links in the linear and nonlinear supply chain optimization models, it is also convenient to monitor the inventory in real time, grasp the detailed information of the goods, improve the level of automation and mechanization of inventory and distribution, and realize a highly modernized logistics operation. Under the construction of the entire IoT architecture, this involves the specification of logistics information representation, the standardization of interfaces between different links, and the input and processing of output file formats from identification tags to label reading devices to transmission and processing [11]. This is beneficial to the information exchange infrastructure between different e-commerce supply chain member companies or different logistics subsystems and to build the largest information platform for information exchange between different individuals, of which the logistics industry is the biggest beneficiary.

The main risks faced by the e-commerce supply chain in the IoT environment are information security risks, reputation risks, and distribution risks in the linear and nonlinear supply chain optimization models. Due to the complexity and diversity of the e-commerce supply chain risk itself, it also requires higher risk assessment methods. Therefore, the selected risk assessment method must be able to deal with some unknown, large fluctuations and cause large losses, and it can give a more accurate assessment of the risks. The framework of the IoT-based e-commerce supply chain model is shown in Figure 1. For supply chain managers, it is necessary to determine the conditions under which the supply chain needs to be adjusted and optimized in order to reduce the maintenance cost of the supply chain while improving the value of the entire supply chain. A complete supply chain will always develop in two directions. First, the

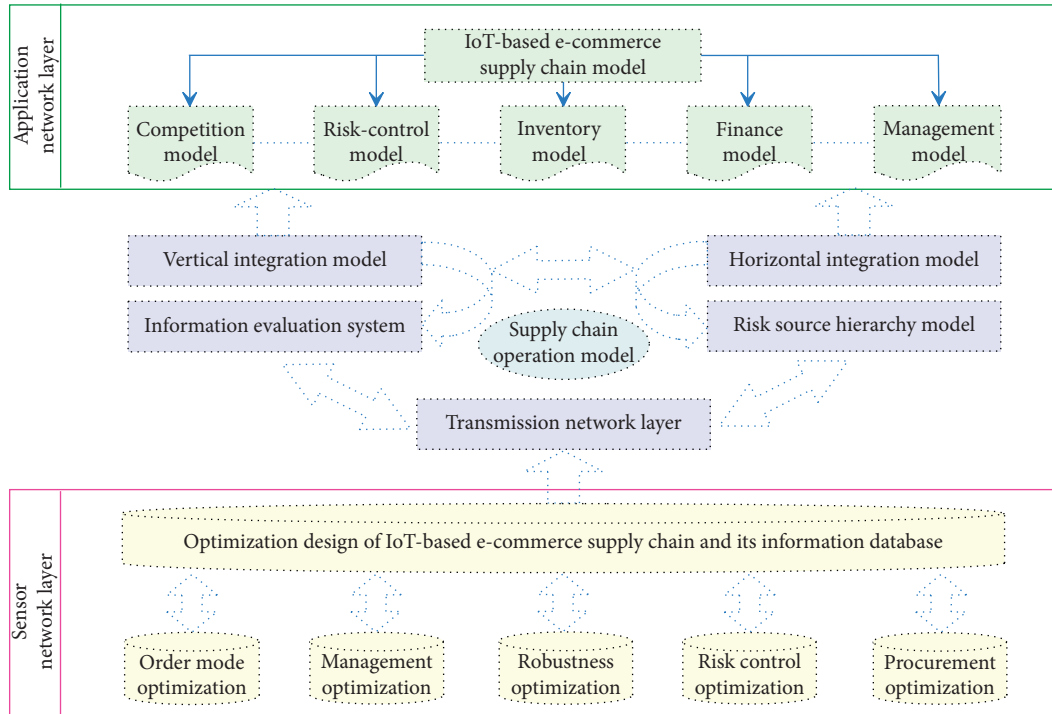


FIGURE 1: Framework of the IoT-based e-commerce supply chain model.

supply chain risks can converge well, and the influence of the supply chain risk factors on each other is reduced to the lowest level at a certain level of risk, and the supply chain maintains stable operation at a certain level of risk through effective management and control of the manager. Second, the supply chain manager's ability to manage supply chain risks is low, causing the supply chain risk level to exceed a critical value, and the supply chain risk cannot effectively converge. The mutual influence of factors is further strengthened, making the risks continue to spread, eventually leading to the collapse of the supply chain [12].

The closed-loop operation mode in the e-commerce supply chain includes the forward supply chain activities of new products from raw materials to consumers, the return of waste products from end users to producers, and the conversion of waste products into remanufactured products by producer activities, product redesign activities for reuse of waste products, environmental monitoring of raw materials, new products, waste products, and green logistics management. Ensuring the accuracy and timeliness of information is the key factor for the success of the closed-loop operation mode of the linear and nonlinear optimization models. The rise of the IoT has provided good development opportunities for its operation and management. Through the IoT platform, all companies in the closed-loop supply chain are connected to form an end-to-end intelligent supply chain logistics distribution service process, which can achieve accurate and effective collaborative distribution based on the real-time status of logistics distribution [13]. The IoT provides an effective information technology support platform for the integration of information and resources in the e-commerce supply chain, tracking and

monitoring of the entire product life cycle. The closed-loop supply chain operation system based on the IoT can implement real-time tracking and supervision of the production, distribution, retail, and recycling of products in the supply chain, enhance management transparency, and enable companies in the e-commerce supply chain to respond quickly to the market demand, improve service levels, and enhance corporate competitiveness.

### 3.2. Risk-Control Model of E-Commerce Supply Chain.

Because many small-volume orders in the traditional supply chain are difficult to access to the factory's mass production plan and subsequent shipment clearance is more troublesome, cross-border e-commerce has provided the possibility to solve these problems. Centralized processing of small orders enables exporters to place orders and produce uniformly, making it easy for factories to adjust product and profit structure. In recent years, as the minimum order quantity is not high, many customers have expressed difficulties in finding good suppliers. Some buyers prefer a supplier with poor quality but a low minimum order quantity, rather than a supplier with a high minimum order quantity but a high quality. Since consumers are not limited by time and space on the e-commerce platform and can browse the information of the products they are interested in, exporters need to pay more attention to the front end of the linear and nonlinear optimization models, actively develop and design products, and improve product quality to attract consumers' attention. The high-quality shops on the online platform will have a higher visibility than the products on the offline exhibition. Large exporters also need

to build brands, promote their brands on cross-border e-commerce platforms, expand product awareness and influence, and establish cross-border e-commerce operations teams to make every node on the line such as customs clearance and logistics. It gives full play to its own offline supply chain mature advantages, thereby forming an integrated advantage of online transactions. Figure 2 shows the relative optimization contribution rate of different optimization terms in the IoT-based e-commerce supply chain model.

The integration of intelligent sensing equipment and the Internet has optimized the basic links of e-commerce supply chain inventory management. The traditional method of manually recording financial data information has lagged behind and cannot match the current market development environment and technology level. For example, the quantitative assessment benchmark is established through corresponding index data. Before the related inventory management model is reformed, the product quality and service levels are difficult to quantify [14]. The integration of the IoT and Internet technologies can establish a sensor network, reduce costs to a certain extent, perform timely distribution, track real-time e-commerce products, reduce the risk of disputes between enterprises and consumers, and protect consumers' basic rights and interests. They improve the service level of the enterprise and increase customer satisfaction, and the e-commerce companies should use empirical data as the basic basis and make corresponding basic management decisions. For example, they should focus on capital investment, sort out changes in market demand, and provide customers with basic value-added services. As a method of supply chain inventory management reform, the use of quantification as an assessment standard is the basic sign of management progress for e-commerce enterprises, and it can be used as a reference for enterprise data demonstration and management reform.

Supply chain finance is a process of systematically optimizing the availability and cost of funds in a corporate ecosystem dominated by core companies. When the sensor of the IoT detects the risk of qualitative changes in the collateral, it timely sends a signal to the information management system and notifies the customer, which facilitates the bank to quickly determine the condition of the collateral and decide whether to recheck the price and compensate for the decline. If the total price drop exceeds the bank's minimum credit line and the borrower has not added pledges or provided new guarantees, immediately announce that the credits will expire prematurely. After communicating with customers, take measures such as auctions to sell pledges and repay loans. Through radio frequency identification technology, the e-commerce and banks can clearly understand and grasp every link of collateral from production and transportation to sales in the linear and nonlinear optimization models. The readers installed in factories, distribution centers, and logistics supervision warehouses can automatically record items throughout the supply of the flow of the chain. From the perspective of the entire integrated supply chain, the transparency of the supply chain and the accuracy of the

information will be maximized, which is exactly the problem of difficult supervision of banks. After the bank has reached a transaction with the on-chain enterprise, the bank delivers the collateral from the enterprise to the cooperative warehouse. The warehouse is responsible for keeping the collateral and regularly checks the value of the collateral with the bank [15].

## 4. Optimization Design of IoT-Based E-Commerce Supply Chain

*4.1. Procurement Robustness Optimization of E-Commerce Supply Chain.* With the support of IoT technology and the e-commerce procurement link, in order to achieve long-term mutual benefit and close cooperation of information sharing among members of the supply chain, it is necessary to continuously improve the management level of suppliers and support a series of supplier management activities. The perfect management system integrates powerful IoT technology, which can produce comprehensive data such as quantity, quality, price, cost, delivery efficiency, and service level and can provide the supplier's supply performance, supply capacity, supply reputation, and technical level. The management efficiency and other indicators better streamline and optimize suppliers and ensure the quality of supply chain procurement [16]. The e-commerce procurement logistics system includes many logistics activities, such as packaging, loading, unloading, transportation, warehousing, handling, and information management from suppliers to producers, and it is an important part of the robust optimization of supply chain procurement (Figure 3). The packaging and transportation links, storage inspection links, and material warehousing links of the e-commerce procurement logistics system need to be fully optimized in accordance with the new IoT environment. Through a comprehensive perception of raw materials, perfecting transportation plans, optimizing logistics routes, and improving procurement models, the scientific management of raw materials storage warehouses, delivery vehicles, and transportation routes is completed to realize flexible and agile procurement.

Delivery management under traditional supply chain management often becomes the last link that affects the overall efficiency of supply chain enterprises. Uncertain transportation information, poor timeliness, and high cargo damage rates are the direct causes of customer satisfaction decline. In e-commerce platforms, on the other hand, upstream companies in the supply chain can release logistics and transportation information in a timely manner to provide downstream companies and consumers with the convenience of querying and tracking. In addition, they can adjust delivery batches or delivery times in a timely manner when market sales conditions change, increasing the number of delivery companies and time flexibility and reducing the risk of corporate material reserves. The enterprise can reasonably arrange the production plan of the product according to the sales situation of the e-commerce platform and can adjust the production capacity and product structure in a timely manner based on the analysis of the data

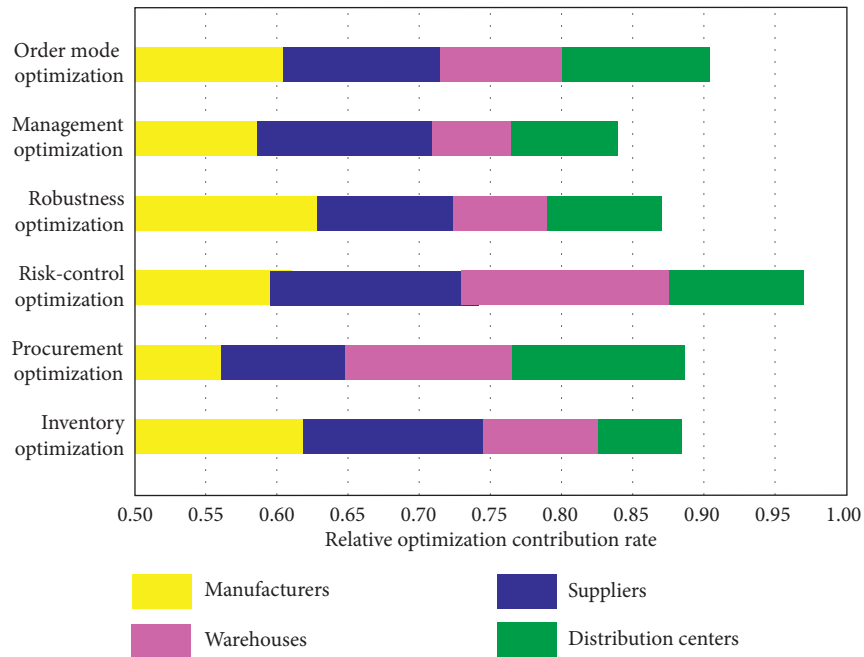


FIGURE 2: Relative optimization contribution rate of different optimization terms in the IoT-based e-commerce supply chain model.

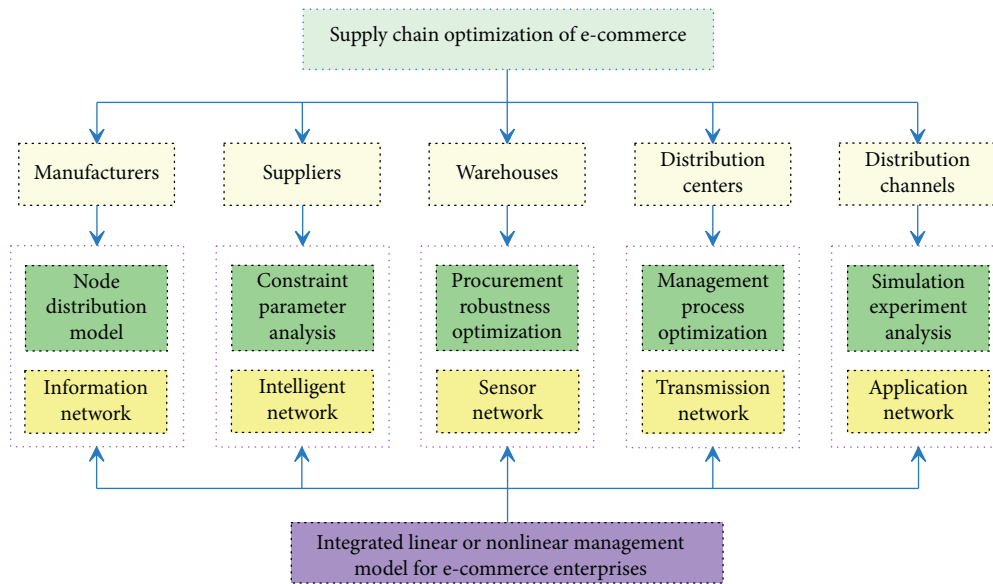


FIGURE 3: Optimization design for IoT-based e-commerce supply chain.

transmitted by the ordering system to further reduce the uncertain factors in production. Through the timely and rapid update of information and integrated processing, it can achieve collaborative management of production and sales, shorten the manufacturing cycle, reduce customer waiting time, and improve customer satisfaction. In terms of delivery and delivery time, real-time query of delivery information allows customers to reasonably arrange the delivery time and ensures the timely supply of goods in the linear and nonlinear optimization models. The customers can also log in to the enterprise's e-commerce platform to query product update information in real time in order to

carry out material procurement planning and inventory planning adjustment, and realize flexible procurement at any time.

Under the support of IoT technology, the connection between e-commerce companies has shown a networked and cross-shaped development. The network's corporate layout will provide a larger financial services industry including banks market space. In addition, the IoT will be used to achieve efficient resource allocation and information sharing, especially the IoT itself, in many areas of e-commerce product research and development and production processes, due to the cross and openness of the network.

With the further promotion and extension of the IoT in various fields such as social production, there will be applications of IoT technologies in everything from raw material production to product quality and logistics distribution. The services provided by this supply chain finance are no longer targeted at a specific customer but are based on the overall situation of the entire IoT system to formulate corresponding development strategies and countermeasures in the linear and nonlinear optimization models. Because the market is constantly changing and developing, only from a global perspective, we can avoid making wrong decisions due to unscientific analysis, which ultimately affect the vital interests of the financial services industry and its customers. Only constructing a three-dimensional and multidimensional IoT platform from the scope and degree will lay a solid technical foundation for the future development of e-commerce supply chain finance [17].

**4.2. Optimization of E-Commerce Supply Chain Management Process.** The supply chain management mode under the e-commerce environment information flow management, capital flow management, and logistics management constitutes the three major parts of supply chain management. The enterprises in the supply chain should consider the various business transactions within the enterprise and with the nodes in the supply chain as an overall functional process. By effectively and transparently coordinating the information flow, capital flow, and logistics in the supply chain, the internal supply chain of the enterprise integrated management with the supply chain outside the enterprise to form an integrated supply chain management system. Because the core enterprises of e-commerce and the upstream suppliers, downstream end customers, banks, and logistics centers can exchange information quickly through the Internet, at the same time, the nodes in the supply chain can also communicate with each other. The IoT-based node distribution of the logistics supply chain with horizontal and vertical integration modes is shown in Figure 4. They overthrow the traditional model and effectively connect the islands of each business node in the supply chain to achieve true business information integration and sharing through the application of e-commerce. At the transaction stage, the business needs to further improve the logistics management link, minimize the time required for logistics in the supply chain, and achieve unified and synchronized logistics management with information flow and capital flow management, thereby establishing a truly powerful and fast response of the supply chain management system.

The IoT technology enables tracking of items flowing in the e-commerce supply chain, and at the same time, it can transmit data to all participants in real time, reducing the phenomenon of information distortion. Fast information transmission speed can greatly reduce the cost of transmission time between roles such as enterprises, merchants, and users. With the development of globalization and the improvement of people's living standards, the demand for personalized customization is getting higher and higher, and

the e-commerce platform has turned to a user-oriented personalized design concept [18]. Each part of the e-commerce supply chain can be associated with users through the network to obtain a wider and more effective interaction and production scheme, so the platform more respects the individual needs of users. Intelligent management of e-commerce supply chain is an integrated management idea and method, which integrates the planning, organization, coordination and control of logistics, information flow, capital flow, value-added flow, business flow, and trading partnership in the linear and nonlinear optimization models. Through supply chain management, the enterprises can improve operational efficiency, improve asset utilization, shorten the total order cycle, reduce transaction costs, enhance customer service, improve response speed, penetrate new markets, obtain higher return on assets, and enhance their competitiveness.

With the widespread application of e-commerce, information and funds are transparently, quickly, and accurately transferred between member nodes of the supply chain through the Internet, which greatly improves the information flow management and capital flow management in supply chain management. Breakthrough to change the traditional scope of raw material procurement, production, sales, and services to meet the new requirements of high quality, high flexibility, and low cost imposed by the market on corporate production management in the new competitive environment is the inevitable requirement of the supply chain management model in the e-commerce environment. The realization of the e-commerce supply chain management process optimization is based on the integration and sharing of information. The efficient operation of the supply chain requires the enterprises of each node to give high-quality information data and fast information flow [19]. The e-commerce supply chain information integration and sharing platform can achieve unified standards for internal and external information, interoperable, seamless docking, integration, analysis, and sharing so that each node company in the supply chain can perceive information in a short time. At the same time, it is convenient for the enterprises of each node to grasp the production progress of the product and the operation of the supply chain in time. Based on the integration of internal and external mass information and the direction of the wind, it promotes the orderly interaction of internal and external management systems and the coordinated development of business.

## 5. Simulation Experiment Analysis

**5.1. Simulation Experiment Design.** According to the IoT-based e-commerce supply chain linear or nonlinear optimization model, the importance order of each influencing factor from least important to most important is circulation sociality, circulation convenience, circulation timeliness, surplus level, circulation quality, and circulation cost.

It is assumed that the selected e-commerce supply chain contains  $n$  optimization processes; the hierarchical single-order calculation of the above six influencing factors by the  $i$ -th process is as follows:

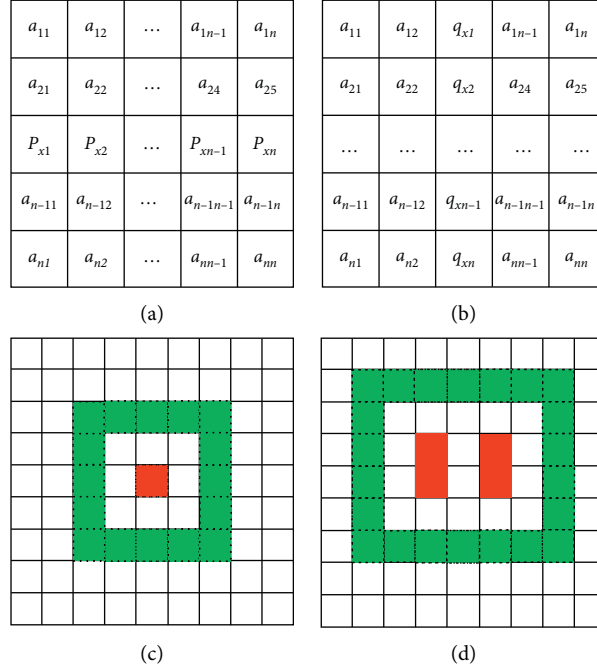


FIGURE 4: IoT-based node distribution of logistics supply chain with horizontal (a, c) and vertical integration modes (b, d).

$$M_{ij} = \sum A_{ij} (i, j = 1, 2, 3, 4, 5, 6). \quad (11)$$

According to the judgment matrix of the supply chain optimization process, it can be clearly obtained that the optimization coefficients ( $O_i$ ) are  $O_1 = 15.34$ ,  $O_2 = 13.66$ ,  $O_3 = 17.19$ ,  $O_4 = 14.03$ ,  $O_5 = 11.35$ , and  $O_6 = 12.89$ .

The optimization weighting factor  $s_i$  is calculated by  $s_i = \sum_{i,j=1}^6 (A_{ij} O_j / M_{ij})$ , and then the results are  $s_1 = 0.54$ ,  $s_2 = 0.23$ ,  $s_3 = 0.48$ ,  $s_4 = 0.56$ ,  $s_5 = 0.44$ , and  $s_6 = 0.29$ .

The optimization quality  $Q_i$  is calculated by  $Q_i = O_i / W_i$ , and then the values were calculated as  $Q_1 = 0.84$ ,  $Q_2 = 0.91$ ,  $Q_3 = 0.75$ ,  $Q_4 = 0.87$ ,  $Q_5 = 0.66$ , and  $Q_6 = 0.93$ .

According to the supply chain before and after the optimization, each level is sorted separately from the particularity of the second-order matrix, and the weight proportion ( $W_1$ ) and priority ( $W_2$ ) of each matrix are calculated as follows. In  $Q_1$ , the proportion of  $W_1$  is 0.31 and the proportion of  $W_2$  is 0.25. In  $Q_2$ , the proportion of  $W_1$  is 0.58 and  $W_2$  ratio is 0.42. In  $Q_3$ ,  $W_1$  ratio is 0.19 and  $W_2$  ratio is 0.85. In  $Q_4$ ,  $W_1$  ratio is 0.56 and  $W_2$  ratio is 0.63. In  $Q_5$ ,  $W_1$  ratio is 0.37 and  $W_2$  ratio is 0.94. In  $Q_6$ , the  $W_1$  ratio is 0.59 and  $W_2$  ratio is 0.66.

Finally, the consistency of the matrix is checked, and the supply chain optimization consistency parameter  $f_n = 0.0537$  under the condition of  $n = 6$ , and the value is less than 0.1, that is, it can pass the consistency check.

**5.2. Result Analysis.** To a certain extent, the structural level of the supply chain can reflect the cost and profit appreciation process of e-commerce products from demand forecasting, production, and processing to completion of sales. The more the layers in the supply chain structure, the more the number of different companies' value chains

linked, the greater the cost, and the worse the economic benefits of the entire value chain in the integrated linear or nonlinear management models (Figure 5). Therefore, reducing the structural level of the e-commerce product supply chain and establishing dynamic cooperation relationships are important ways for enterprises to reduce bad costs and improve operating results. The optimization efficiency varies from 0.96 to 0.50 when the weighting factor equals 0.30 in the linear model while it varies from 0.88 to 0.61 in the nonlinear model. The low entry cost of virtual markets, as well as the uneven distribution of market information reformed by information overload and dynamic information flow, makes the information asymmetry between economic entities continue to exist within a certain range. In most cases, this sedimentary value is mainly manifested as the cost of invalid logistics and continues to be a defect in the overall economic benefits of the e-commerce product supply chain. All members in the supply chain should be a unified and harmonious division of labor and cooperation, while the characteristics of the traditional supply chain structure are relatively complex, the members are relatively stable, and the interdependence is strong and harmony between members of the supply chain in the linear and nonlinear optimization models.

Different from the previous concept of the operation level of e-commerce supply chain, as an intermediary variable, the mechanism of IoT application affecting enterprise performance is studied. Supply chain integration is used as an intermediary variable to study the mechanism of IoT application affecting enterprise performance. The intermediary role of supplier integration and customer integration can affect business performance. The IoT applications can affect e-commerce business performance through the full mediation role of supplier integration but only through the

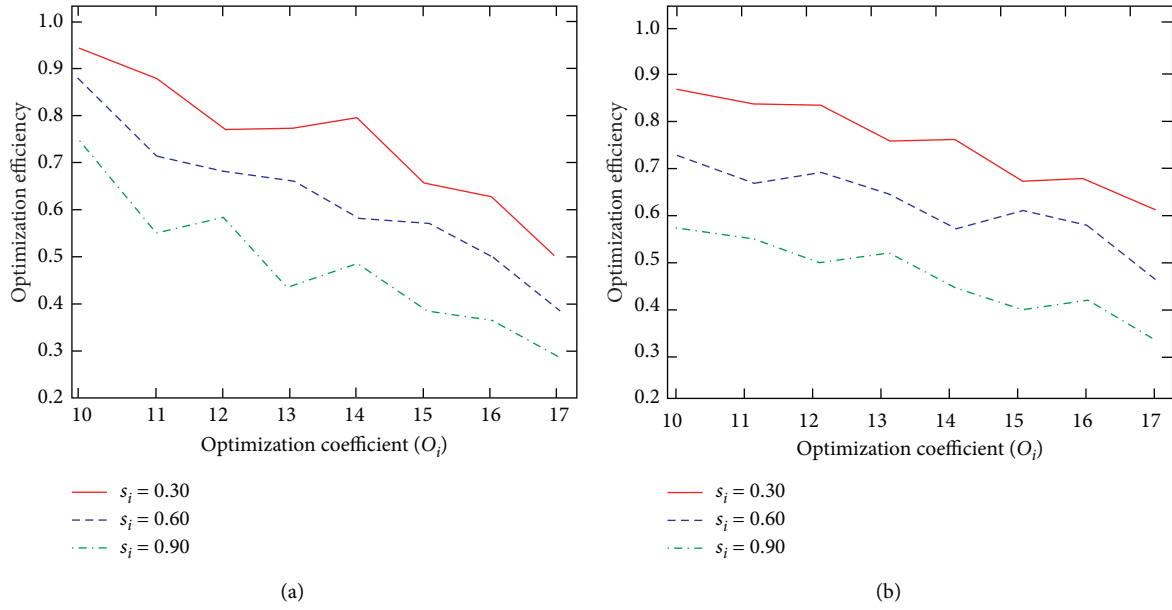


FIGURE 5: Relationship between optimization efficiency and coefficient with different weighting factors ( $s_i$ ) in the linear (a) and nonlinear (b) management model.

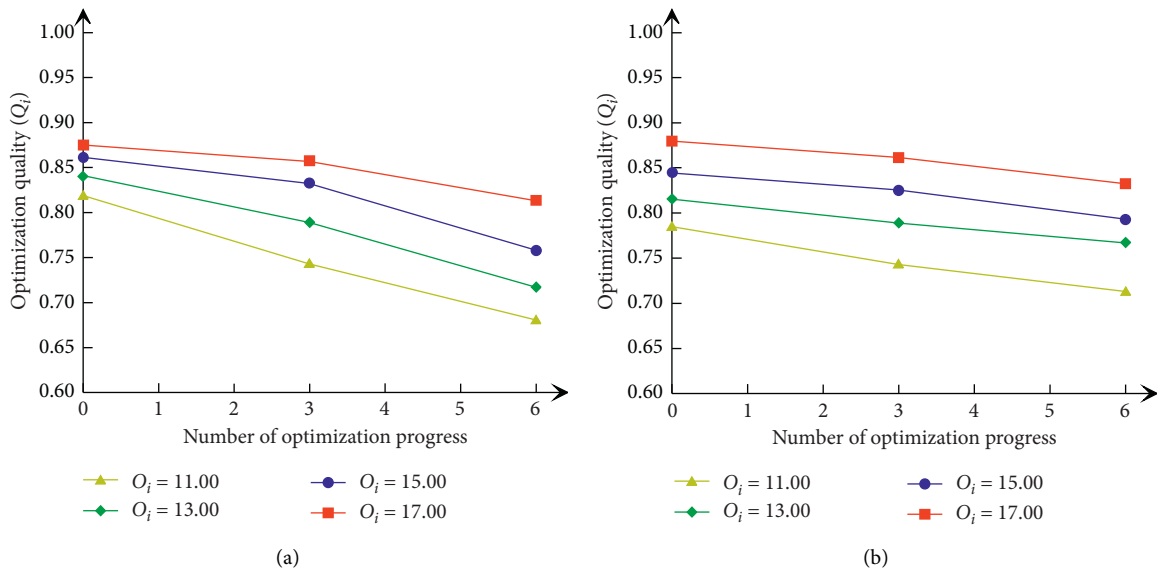


FIGURE 6: Relationship between optimization quality and the number of optimization progress with different optimization coefficients ( $O_i$ ) in the linear (a) and nonlinear management model (b).

partial mediation role of customer integration. The e-commerce can not only link the IoT with existing processes, products, or services but also the IoT with new processes, products, or services. In the process of IoT application, e-commerce companies should pay attention to the level of supply chain integration, especially supplier integration, because the two types of IoT applications can improve the performance of enterprises through the full mediation of supplier integration and the partial mediation of customer integration (Figure 6). The optimization quality changes from 0.87 to 0.84 when optimization coefficient

equals 11.00 in the linear model while it changes from 0.87 to 0.85 in the nonlinear model. Therefore, the enterprises should increase investment in supplier and customer relationships, establish mechanisms for sharing information with suppliers and customers, actively coordinate conflicts with suppliers and customers, and promote joint resolution of problems with suppliers and customers in the linear and nonlinear optimization models, thereby enhancing the company's relationship with the level of supplier and customer integration and then the performance impact of IoT applications [20].



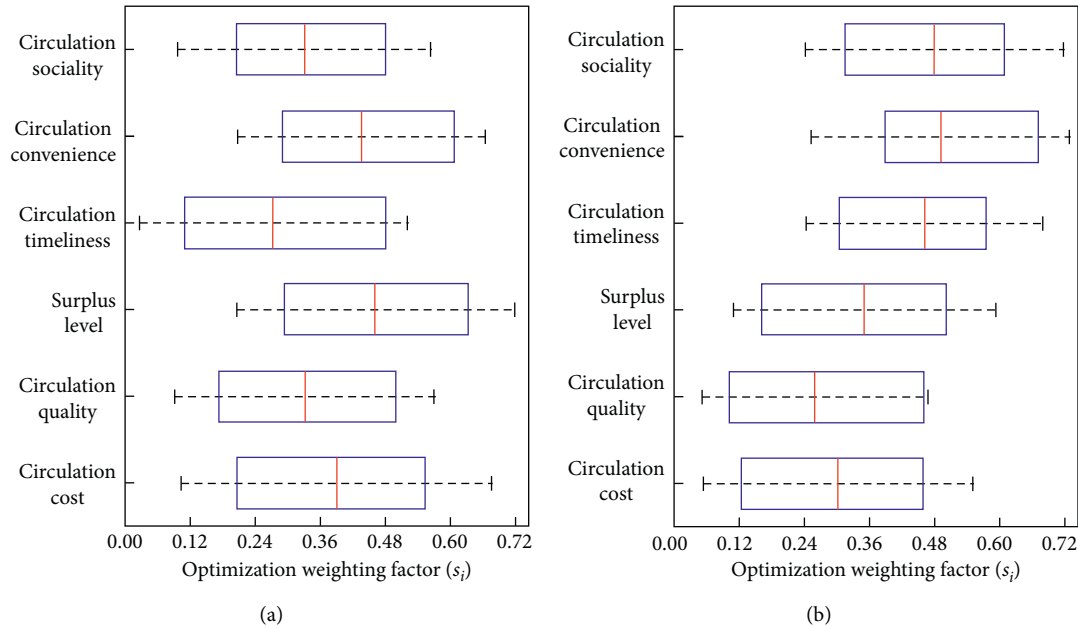


FIGURE 7: Optimization weighting factors ( $s_i$ ) of influential factors in IoT-based e-commerce supply chain with horizontal (a) and vertical (b) integration modes.

Cost optimization is the main way for enterprises to win market competition. The implementation of e-commerce-based procurement supply chain management can effectively reduce the cost of global procurement for enterprises and make the procurement supply chain cost optimized. The optimization weighting factors of influential factors in IoT-based e-commerce supply chain with horizontal and vertical integration modes are shown in Figure 7. The procurement management system implements procurement management based on the supplier's credit points to support enterprises to realize the supply control process in the supply chain environment. Through the timely and rapid update of information and integrated processing, it can achieve collaborative management of production and sales, shorten the manufacturing cycle, reduce customer waiting time, and improve customer satisfaction. The inventory management system uses the supplier's joint inventory management strategy to share the e-commerce enterprise's inventory information with the supplier to achieve supply integrated optimization of chain inventory and service level [21]. The management efficiency of the supply chain depends on the coordination between enterprises through information sharing. The basis of information sharing is to establish an integrated management information system to increase the supply chain members to obtain information timeliness and visibility to improve demand forecast accuracy [22]. In the IoT environment, e-commerce can give full play to the synergies of products, processes, and information, reduce loops and cycles, effectively avoid problems caused by information asymmetry through system construction, improve technology, and make the entire system and management construction of the process cover all aspects of information sharing requirements, which further improves the effect of application integration and standardizes and collects information in the supply chain.

## 6. Conclusions

This article expounded the research status and significance of e-commerce supply chain optimization, elaborated the development background, current status, and future challenges of the IoT technology, introduced the methods and principles of the node distribution model and constraint parameter analysis of logistics supply chain, constructed an IoT-based e-commerce supply chain model including competition and risk-control modules, performed the design of IoT-based e-commerce supply chain optimization, and discussed the solutions to procurement robustness optimization and management process optimization. The optimization efficiency varies from 0.96 to 0.50 when the weighting factor equals 0.30 in the linear model while it varies from 0.88 to 0.61 in the nonlinear model. The optimization quality changes from 0.87 to 0.84 when the optimization coefficient equals 11.00 in the linear model while it changes from 0.87 to 0.85 in the nonlinear model. With the support of IoT technology, the e-commerce procurement link requires long-term mutual benefit and close cooperation of information sharing among members of the supply chain, which is necessary to continuously improve the management level of suppliers and support a series of supplier management activities. The final simulation analysis showed that the proposed optimization model can effectively predict e-commerce product information in each stage of the supply chain; the IoT-based e-commerce supply chain optimization can enable each node company in the supply chain to improve the quality of information transmission and reduce the risk of information asymmetry in the process of e-commerce exchanges and can quickly and easily respond to the sudden changes of market and environment, thereby promoting them to cooperate with each other for

establishing a supply chain network structure with closer business relationships, smoother logistics transportation, more reasonable benefit sharing, and stronger market competitiveness. The study results of this paper provide a certain reference for further research studies on the e-commerce supply chain optimization in the IoT environment.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The author declares no known conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This work was supported by the Guangdong Province Natural Science Fund Project: Construction and Innovation of E-commerce Models by empirical study (no. S2013010011522) and Zhuhai Science and Technology Project: Research on the Development of Internet of Things industry in Zhuhai (no. 2011B030102006).

## References

- [1] M. Abdel-Basset, G. Manogaran, and M. Mohamed, "Internet of things (IoT) and its impact on supply chain: a framework for building smart, secure and efficient systems," *Future Generation Computer Systems*, vol. 86, pp. 614–628, 2018.
- [2] E. Manavalan and K. Jayakrishna, "A review of internet of things (IoT) embedded sustainable supply chain for industry 4.0 requirements," *Computers and Industrial Engineering*, vol. 127, pp. 925–953, 2019.
- [3] F. Caro and R. Sadr, "The internet of things (IoT) in retail: bridging supply and demand," *Business Horizons*, vol. 62, no. 1, pp. 47–54, 2019.
- [4] A. Haddud, A. R. DeSouza, A. Khare, and H. Lee, "Examining potential benefits and challenges associated with the internet of things integration in supply chains," *Journal of Manufacturing Technology Management*, vol. 28, no. 8, pp. 1055–1085, 2017.
- [5] Y. Zhang and J. Wen, "The IoT electric business model: using blockchain technology for the internet of things," *Peer-to-Peer Networking and Applications*, vol. 10, no. 4, pp. 983–994, 2017.
- [6] O. Elijah, T. A. Rahman, I. Orikumhi, C. Y. Leow, and M. N. Hindia, "An overview of internet of things (IoT) and data analytics in agriculture: benefits and challenges," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3758–3773, 2018.
- [7] X. Shang, R. Zhang, X. Zhu, and Q. Zhou, "Design theory, modelling and the application for the internet of things service," *Enterprise Information Systems*, vol. 10, no. 3, pp. 249–267, 2016.
- [8] S. Liu, Y. Zhang, Y. Liu, L. Wang, and X. V. Wang, "An "internet of things" enabled dynamic optimization method for smart vehicles and logistics tasks," *Journal of Cleaner Production*, vol. 215, pp. 806–820, 2019.
- [9] F. Javed, M. K. Afzal, M. Sharif, and B. S. Kim, "Internet of things (IoT) operating Systems support, networking technologies, applications, and challenges: a comparative review," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2062–2100, 2018.
- [10] P. Liu and S. P. Yi, "Pricing policies of green supply chain considering targeted advertising and product green degree in the big data environment," *Journal of Cleaner Production*, vol. 164, pp. 1614–1622, 2017.
- [11] L. F. Hsu, "E-commerce model based on the internet of things," *Advanced Science Letters*, vol. 22, no. 10, pp. 3089–3091, 2016.
- [12] S. S. Kamble, A. Gunasekaran, H. Parekh, and S. Joshi, "Modeling the internet of things adoption barriers in food retail supply chains," *Journal of Retailing and Consumer Services*, vol. 48, pp. 154–168, 2019.
- [13] A. Musa and A. A. A. Dabo, "A review of RFID in supply chain management: 2000–2015," *Global Journal of Flexible Systems Management*, vol. 17, no. 2, pp. 189–228, 2016.
- [14] B. D. Chung, S. I. Kim, and J. S. Lee, "Dynamic supply chain design and operations plan for connected smart factories with additive manufacturing," *Applied Sciences*, vol. 8, no. 4, p. 583, 2018.
- [15] C. Liu, Y. Xiao, V. Javangula, Q. Hu, S. Wang, and X. Cheng, "NormaChain: a blockchain-based normalized autonomous transaction settlement system for IoT-based E-commerce," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4680–4693, 2018.
- [16] S. Tiwari, H. M. Wee, and Y. Daryanto, "Big data analytics in supply chain management between 2010 and 2016: insights to industries," *Computers & Industrial Engineering*, vol. 115, pp. 319–330, 2018.
- [17] A. Boiko, V. Shendryk, and O. Boiko, "Information systems for supply chain management: uncertainties, risks and cyber security," *Procedia Computer Science*, vol. 149, pp. 65–70, 2019.
- [18] Q. Tao, C. Gu, Z. Wang, J. Rocchio, W. Hu, and X. Yu, "Big data driven agricultural products supply chain management: a trustworthy scheduling optimization approach," *IEEE Access*, vol. 6, pp. 49990–50002, 2018.
- [19] Y. Zhang, "Application of improved BP neural network based on e-commerce supply chain network data in the forecast of aquatic product export volume," *Cognitive Systems Research*, vol. 57, pp. 228–235, 2019.
- [20] M. Zhang, G. Q. Huang, S. X. Xu, and Z. Zhao, "Optimization based transportation service trading in B2B e-commerce logistics," *Journal of Intelligent Manufacturing*, vol. 30, no. 7, pp. 2603–2619, 2019.
- [21] A. Saif and S. Elhedhli, "Cold supply chain design with environmental considerations: a simulation-optimization approach," *European Journal of Operational Research*, vol. 251, no. 1, pp. 274–287, 2016.
- [22] J. Huang and J. Song, "Optimal inventory control with sequential online auction in agriculture supply chain: an agent-based simulation optimisation," *International Journal of Production Research*, vol. 56, no. 6, pp. 2322–2338, 2018.

## Research Article

# A Three-Dimensional Complex Measurement Model-Based Avionic Radio-Frequency Power Source Health Assessment Method

Lin Huo <sup>1,2</sup>, Shiqi Li,<sup>1</sup> Simiao Fei,<sup>3</sup> and Chuan Lyu<sup>4</sup>

<sup>1</sup>School of Safety Engineering, Shenyang Aerospace University, Shenyang 110135, China

<sup>2</sup>Liaoning Key Laboratory of Aircraft Safety and Airworthiness, Shenyang, China

<sup>3</sup>Shenyang Aircraft Design Institute, Shenyang 110035, China

<sup>4</sup>Department of Reliability and System Engineering, Beihang University, Beijing 100191, China

Correspondence should be addressed to Lin Huo; huolin@buaa.edu.cn

Received 11 July 2020; Revised 5 September 2020; Accepted 18 September 2020; Published 19 October 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Lin Huo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Health assessment is an important part of PHM technology, which is crucial to product state monitoring and management. With the complexity of product structure and diversification of functions, the product health presents uncertainty under the complex influence of multiple factors. Nowadays, the general health assessment method is relatively simple and only considers the product function completeness, and the assessment effect accuracy still needs to be strengthened in the actual working environment. Most assessment parameters and assessment methods are selected ignoring the impact of environmental changes and operating time performance. Therefore, a new health measurement method for the avionic radio-frequency power source based on a three-dimensional complex measurement model is proposed. Firstly, the product health definition is proposed, and the health connotation is analyzed from three-dimensional aspects containing functional integrity, environmental adaptability, and temporal sustainability. According to the function and structure characteristics of radio-frequency power source, the health characterization parameters are then obtained and the health assessment parameter system is established. Finally, the three-dimensional complex measurement model is given, and the RF power source health assessment is carried out comprehensively from integrity, stability, and reliability aspects. The three-dimensional health assessment method provides a new way to solve the uncertainty of product health state under multiple factor effects, which is conducive to the targeted optimization management strategy.

## 1. Introduction

Prognostic and health management (PHM) technology refers to the use of advanced sensor integration, with the aid of various algorithms and intelligent models to diagnose, predict, and monitor the product health state and reasonably manage the product operation and maintenance so as to obtain the best product health state with the minimum investment. It has evolved from reliability analysis, testability design, fault analysis, and system health management [1]. With the economy development, people's living standards are constantly improving, and the safety awareness is also constantly improving. PHM technology is more and

more widely used, from the initial fighter field to the present civil aviation, aerospace vehicles, computer systems, large-scale mechanical equipment, nuclear power plants, complex product systems, and other fields.

Nowadays, the PHM technology system is developing towards a more intelligent, more precise, more comprehensive, and more convenient direction. Health management is an important part of PHM technology. Its purpose is to manage evaluation data, take proactive measures to monitor the complex product health state, and predict the performance change trend, failure time, and remaining useful life so as to take necessary measures to alleviate the complex product performance degradation [2]. As a state of

product, health has no clear definition, so the product health assessment method is different according to different people's understanding of health. According to the current research situation at home and abroad, health assessment methods are summarized into the following three categories: evaluation methods based on traditional methods, evaluation methods based on mathematical and physical models, and evaluation methods based on data.

Traditional health assessment methods include analytic hierarchy process (AHP), fuzzy theory and grey clustering theory, and so on. Wang et al. used the evaluation model combining fuzzy analytic hierarchy process, fuzzy preference programming, and order performance similarity ideal solution technology to evaluate the aeroengine health state [3]. Ren et al. combined grey clustering theory with rough set theory to form a fast and accurate fault classification decision-making method for numerical control machine tool health assessment [4]. Zhu and Song used cross fuzzy entropy to measure the similarity between test samples and normal samples and used similarity to evaluate the performance degradation state of bearings so as to complete the health state assessment of bearings [5]. Wang et al. combined the analytic hierarchy process (AHP) with variable weight processing to realize the comprehensive health assessment of diesel generator set operation state [6].

Health assessment methods based on mathematical and physical models include the application of Markov chain and so on. Allen et al. proposed a machine fault detection method using the hidden Markov model in process control, and the effectiveness of the method was illustrated by numerical calculation [7]. Bin et al. proposed a method combining Mahalanobis distance and histogram method to construct a complex system health assessment method for the comprehensive evaluation of system health state [8]. Liao et al. applied a new and flexible system or component prediction framework based on high-order hidden semi-Markov model (HOHSM) to assess the health state of NASA turbofan engine [9].

With the development of science and technology, artificial intelligence has become a hot spot of learning and is widely used in various fields; of course, the field of health assessment is no exception. For example, health assessment methods based on data are widely used. Samanta proposed the performance of artificial neural network and support vector machine in gear fault detection [10]. Xu et al. proposed a fault diagnosis method based on the fusion of neural network and D-S evidence theory to diagnose turbine fault [11]. Liu and Zio established a dynamic evaluation and fault prediction model of interconnected component system with noise monitoring data by using parallel Monte Carlo simulation and recursive Bayesian method [12]. Qin et al. developed a GIS equipment operation state analysis model based on machine learning algorithm and evaluated GIS equipment state according to the severity of PD [13]. Oluwasegun and Jung proposed a general framework based on discrete Bayesian network (BN), which is particularly suitable for decision fusion of heterogeneous prediction methods. BN parameters were calculated according to the fixed prognosis target. The validity of the proposed

prediction method based on decision fusion is proved by estimating the remaining useful life of turbofan engine [14]. Kexiong established a health assessment model of missile inertial navigation platform system by combining expert system with data driver according to the characteristics of a missile inertial navigation platform system [15]. Zhang et al. proposed a bearing health assessment method based on Hilbert transform envelope analysis and cluster analysis [16].

Most health assessment methods nowadays are based on the existing traditional assessment methods or modified by one or several methods. The assessment results are relatively simple, and the assessment effect accuracy still needs to be strengthened in the actual working environment. Most parameters are selected according to the product characteristics, ignoring the impact of environmental changes and operating time on product performance.

In this paper, a new multidimensional health assessment method is proposed to evaluate the health of a certain avionic radio-frequency (RF) power source. The model is built from three dimensions of functional integrity, environmental adaptability, and temporal sustainability. The health of the equipment is described comprehensively from three aspects of integrity, stability, and reliability so that the health level is no longer one plane point line, but a number of three-dimensional verticals. It provides a new idea to solve the uncertainty of product health state under the effects of multiple factors, which is conducive to the formulation of health optimization management strategy.

## 2. Health Connotation and Three-Dimensional Complex Measurement Model

Combined with the modern product characteristics, such as diversified functions, changeable use environment, and complex structure levels, product health is defined as the product ability degree/state to continuously respond to the environment and complete the specified tasks. When the product function is intact, only with good sustainability and stability can the product be in the best state. Otherwise, it may reduce or damage the function and cause failure. Therefore, a product in a healthy state should not only have good functions to ensure its ability to perform tasks but also maintain the stability of function, internal structure, and organization for a long time and have certain adaptability to environmental effects. Therefore, the meaning of health is diverse and extensive. The composition of product health should include functional integrity, temporal sustainability, and environmental adaptability.

In general, the changes of product health state are functional failure, sustainability, and the decline of adaptability to the external world. It has two interrelated characteristics: the increase of failure rate with time and the decline of adaptability to environmental changes. Therefore, focusing on three aspects of product health, a product health three-dimensional spatial measurement model is proposed, and the parametric modeling method is studied from the three dimensions of function, environment, and time. Among them, functional integrity is more focused on whether the function of the product can be realized,

temporal sustainability can be considered from the perspective of reliability and failure rate, and the essence of environmental adaptability is to investigate the stability of product performance in a certain environment. Therefore, the research of RF power source health measurement is to study whether the function of RF power source is complete, whether it has certain anti-interference stability, and how long it can work continuously without fault.

**2.1. Functional Integrity Dimension.** In these three aspects, the functional completeness and integrity is the basis of health. The most basic requirement of a healthy product is to be able to realize all its functions normally. The so-called function refers to the ability that a product should have when it runs normally. It is an abstract description of the tasks that a product can accomplish. It describes the inherent health capability of a product from a static state. The functional integrity of a product directly shows whether the product function can be fully realized. Therefore, the functional integrity  $C$  of the product is taken as the measurement index of its functional integrity, and the importance of each function is taken as the weight to measure the conformity between the functions and the complete functions of the product at the present stage, that is, the functional coverage rate, as shown in Figure 1. The value range of health degree is  $[0, 1]$ . When the research object meets all the design requirements (such as functional integrity, environmental adaptability, and temporal sustainability), the health degree is 1, which indicates a technical state that fully meets the design requirements. When the product fails to reach the required health level, the health degree is 0, and the product is in a complete failure state. The intermediate state is a subhealthy state with some defects. In practical application, we understand “the health degree state 1” as “the health design state value 1”.

Suppose that the product has  $m$  functions  $F_i$ , and the relationship between the functions is clear; let the importance weight of each function  $\omega_i \in [0, 1]$ ,  $i = 1, 2, \dots, m$ . Because the relationship between product functions directly affects the realization of the whole function, the expression of function integrity is different according to different functional relationships.

- (1) When the function relationship is in series, the functions are interrelated and indispensable. As long as one function fails, the realization of the function will be affected and the whole function cannot be realized. The product functional integrity is expressed as

$$C = \prod_{i=1}^m \|F_i\|, \quad (1)$$

$$\|F_i\| = \begin{cases} 1, & F_i \text{ function implementation,} \\ 0, & F_i \text{ function not implemented,} \end{cases} \quad i = 1, 2, \dots, m.$$

- (2) When the function relationship is in parallel, each function relationship is independent. As long as one function can be realized, it will not affect the realization of the whole function. However, the failure of multiple functions will directly lead to the decline of the product health state and reliability. It can be seen in product with reconfiguration and spare parts structure. The function integrity of the product is expressed as

$$C = \sum_{i=1}^m \omega_i \|F_i\|, \quad (2)$$

$$\|F_i\| = \begin{cases} 1, & F_i \text{ function implementation,} \\ 0, & F_i \text{ function not implemented,} \end{cases} \quad i = 1, 2, \dots, m.$$

In general, the functions of a product are not only in series or in parallel. In this case, according to the specific product function relationship or task reliability block diagram, the correlation of the functions should be analyzed, and the functional integrity expression can be obtained. However, when the product functional structure is more complex, especially for some complex electromechanical product, it is impossible to give a clear functional relationship. Then, the characterization parameters describing the product function can be used as variables to measure the conformity between the function and the functional integrity of the product at this stage, and the overall functional integrity of the product can be obtained.

Assuming that the characterization factor of product function is  $f$ , the expression form of product functional integrity  $C$  is

$$\forall C \in [0, 1], \quad \text{s. t. } X_F \vartheta C, \quad (3)$$

$$C = \vartheta[X_F - T], \quad (4)$$

where  $X_F$  is the parameter representing the functional state of the product;  $T$  is the standard parameter with full functionality corresponding to parameter  $X_F$  and  $\vartheta(\cdot)$  is the functional deviation function;  $\vartheta: X_F | R \rightarrow h_f | [0, 1]$  indicates that the mapping  $\vartheta$  projects the functional characterization parameter  $X_F$  into a real number of  $C \in [0, 1]$ .

The projection of  $X_F$  is a real number of  $C \in [0, 1]$ .

In some complex conditions, when the product monitoring data are difficult to obtain, it is impossible to measure the environmental adaptability or temporal sustainability, and the product health can also be simplified as functional integrity assessment.

**2.2. Environmental Adaptability Dimension.** Health is not just the synonym of good function. Healthy product can maintain the stability of function and structure for a long time and can realize all the predetermined functions, performance, or the ability not to be destroyed under the interference of external environment, that is, it has the ability

to adapt to the environment. Currently, healthy subjects with normal parameters are likely to have fatal problems once they are disturbed or changed due to lack of environmental adaptability.

Product environmental adaptability refers to the product ability to realize all its functions (properties) and not be damaged under the influence of various environments expected to be encountered in its life cycle, which is an important product quality characteristic [17]. It is inseparable from health. Environmental adaptability reflects the adaptability and stability of product function and performance to environmental changes, while functional integrity is to investigate whether the product can meet certain functional standards. In other words, the better the anti-interference and stability of the product under certain environmental conditions, the stronger the ability to maintain its structure and function and the stronger the environmental adaptability of the product. Therefore, product environmental adaptability is essentially to investigate the stability of the product health characterization parameters under the external environmental interference, that is, the noise cancelling ability.

The signal-to-noise ratio (SNR) of communication theory is introduced as a parameter index of product environmental adaptability. The signal-to-noise ratio (SNR) is the ratio of the effective part to the invalid part of the research object; it is commonly used in signal processing, speech recognition, image processing, and quality control [18–20]. Here, the output data of product health characterization parameters can be regarded as useful signal part, while the noise caused by environmental interference is useless signal part. The higher the SNR, the better the adaptability and anti-interference ability of the product to the external environment and the better the environmental adaptability of the product, and vice versa.

The signal-to-noise ratio (SNR), as a parameter to measure the ratio of useful signal to noise, originally came from the field of communication [18], which refers to the ratio of signal power/amplitude to total noise power/amplitude. The calculation formula is as follows:

$$\text{SNR} = \frac{P_S}{P_N} = \left( \frac{A_S}{A_N} \right)^2, \quad (5)$$

where  $P_S$  and  $A_S$  are signal power and amplitude and  $P_N$  and  $A_N$  are noise power and amplitude. In order to get the signal-to-noise ratio more intuitively, we need to convert the time-domain signal to the frequency domain, so we can calculate the signal-to-noise ratio  $S_N$  through Fourier transform, as shown in Figure 2.

$$S_N = \text{SNR} = \text{signal peak} - \text{noise floor} - 10 \lg N \text{ (dB)}, \quad (6)$$

where signal peak is the peak value of the signal; noise floor is the base of noise; and  $N$  is the number of samples.

**2.3. Temporal Sustainability Dimension.** The product health is closely related to its complete function, good stability, and sustainability, which is manifested in functional failure, increased failure rate with time, and decreased adaptability

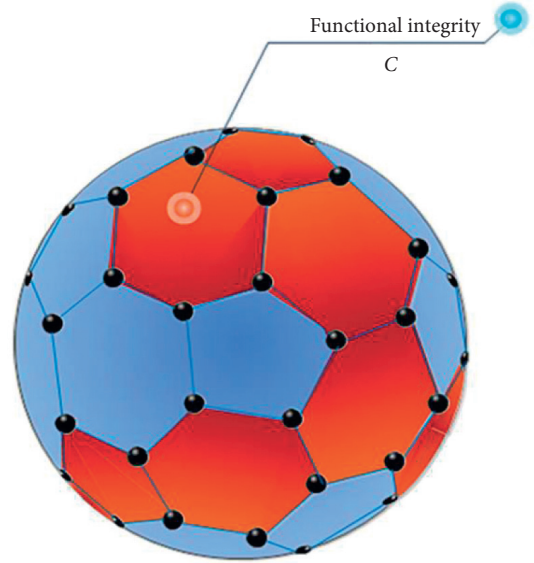


FIGURE 1: Schematic diagram of functional integrity.

to environmental changes. Good temporal sustainability can greatly increase the reliable time between failure so as to maintain the stability of its function and structure. Therefore, the product temporal sustainability is studied from the perspective of reliability.

In the theory of reliability, time is the core of reliability. Generally speaking, the longer the normal work (time between failure) is, the higher the reliability is. The definition of basic reliability in GJB451-90 refers to the product non-failure duration or probability under required conditions [21]. We hope that the product can continue to have the required capability for a period of time. The reliability function  $R(t)$  precisely describes the possibility of the product to perform and maintain its function in a certain period of time  $[0, t]$ , as shown in Figure 3.

As shown in Figure 3, with the increase of operating or storage time, the reliability of the product decreases and the failure rate increases, which means that the continuous operation time of product reliability decreases and the temporal sustainability decreases. Therefore, from the perspective of time, reliability can fully reflect the continuous working ability. Therefore, this paper takes reliability as the measurement parameter index of temporal sustainability.

The higher the reliability  $R(t)$  of the product is, the smaller the failure rate is, and the higher the probability that the product can maintain the required function in time  $[0, t]$ , the longer the sustainable working time is, that is, the better the product temporal sustainability; otherwise, the lower the reliability is, the shorter the product sustainable working time is and the worse the temporal sustainability is.

Therefore, through the use of functional integrity, signal-to-noise ratio, and reliability, a three-dimensional spatial model of RF power health is established, which comprehensively describes the health of RF power source from three aspects of integrity, stability, and reliability.

As shown in Figure 4, the health expression of the three-dimensional space of the RF power source is recorded as

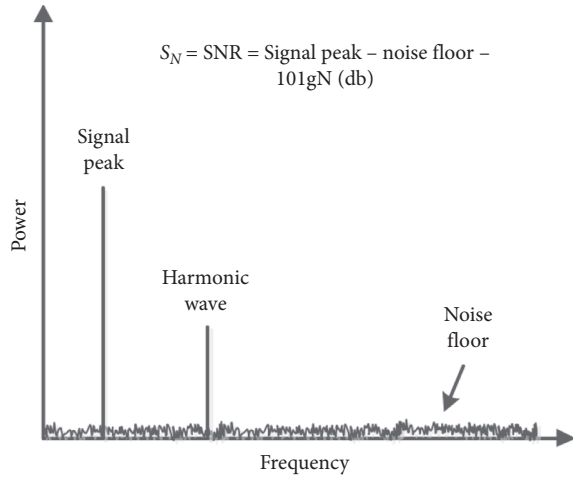


FIGURE 2: Schematic diagram of SNR calculation based on Fourier transform.

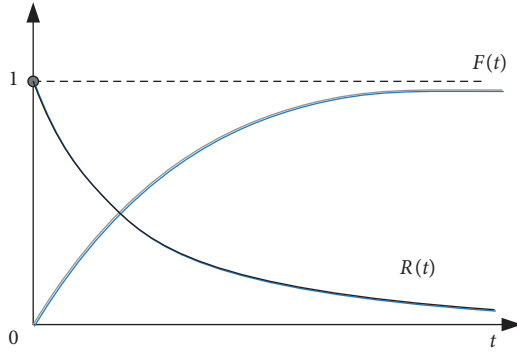


FIGURE 3: Reliability function and failure distribution function curve.

$$H_{\Delta} = (C, S_N, R), \quad (7)$$

where  $C$  is functional integrity;  $S_N$  is the signal-to-noise ratio; and  $R$  is the reliability.

$$H_i = \begin{cases} \{\text{unhealthy}\}, & \{[0, C_{\min}) \cup (-\infty, S_{\min}) \cup [0, R_{\min})\}, \\ \{\text{subhealthy}\}, & 1 - \{[0, C_{\min}) \cup (-\infty, S_{\min}) \cup [0, R_{\min})\} - \{[C_{\min}, +\infty) \cap [S_{\max}, +\infty) \cap [R_{\max}, +\infty)\}, \\ \{\text{healthy}\}, & \{[C_{\min}, +\infty) \cap [S_{\max}, +\infty) \cap [R_{\max}, +\infty)\}. \end{cases} \quad (9)$$

When the product can meet all the ability requirements of the three dimensions, the product belongs to the healthy state, that is, all of them reach or exceed the health threshold of the three dimensions  $[C_{\max}, +\infty) \cap [S_{\max}, +\infty) \cap [R_{\max}, +\infty)$ —green space area; when at least one of the three dimensions does not reach the health limit and these three dimensions are above the abnormal threshold, then the product health is in the subhealthy state—yellow space area. As long as a dimension drops below the threshold of the exception, i.e.,  $[0, C_{\min}) \cup (-\infty, S_{\min}) \cup [0, R_{\min})$ , the product is in a dangerous and unhealthy state—red space

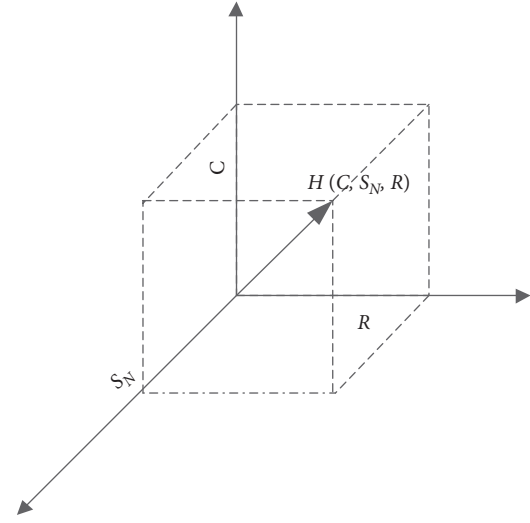


FIGURE 4: Three-dimensional health measurement model.

If the threshold range of the three dimensions of health is determined, respectively, the three levels can be divided into normal, dangerous, and abnormal (or failure).

$$C = \begin{cases} [0, C_{\min}), & \{\text{abnormal}\}, \\ [C_{\min}, C_{\max}), & \{\text{dangerous}\}, \\ [C_{\max}, +\infty), & \{\text{normal}\}, \end{cases} \quad (8)$$

$$S_N = \begin{cases} (-\infty, S_{\min}), & \{\text{abnormal}\}, \\ [S_{\min}, S_{\max}), & \{\text{dangerous}\}, \\ [S_{\max}, +\infty), & \{\text{normal}\}, \end{cases}$$

$$R = \begin{cases} [0, R_{\min}), & \{\text{abnormal}\}, \\ [R_{\min}, R_{\max}), & \{\text{dangerous}\}, \\ [R_{\max}, +\infty), & \{\text{normal}\}. \end{cases}$$

Therefore, it can be divided into three-dimensional health level space, as shown in Figure 5. The health degree of the product can be obtained as follows:

area; until the three dimensions are reduced to zero, the product develops to the state of complete failure.

### 3. Construction of Index System for Health Characterization Parameters of RF Power Source

The typical power module of a certain equipment RF channel system is selected as the research object. It is known that the power module, antenna interface module, frequency

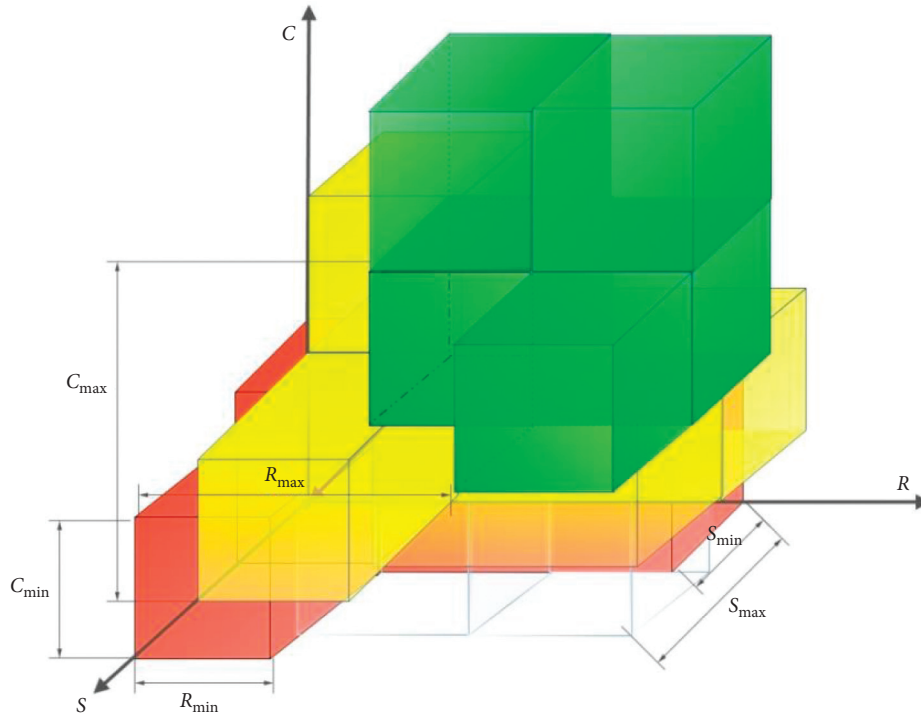


FIGURE 5: Three-dimensional spatial range of health degree.



FIGURE 6: A step-down power source product.

conversion channel module, digital interface, and processing module constitute the whole RF channel system. It is the total power source of each component module in the system. It completes the functions of signal receiving and sending, signal channel switching, signal frequency conversion processing, and signal preprocessing, commonly used in aerospace equipment such as aircraft.

The power module is a typical step-down power source product, as shown in Figure 6. It includes DC/DC converter and filter, which converts DC 270 V to DC 28 V and performs filtering output and provides overheating protection, overvoltage protection, and short-circuit protection.

**3.1. Construction of RF Power Health Characterization.** The key health characterization factors of RF power source were determined and screened step by step. The specific steps are as follows.

**3.1.1. Defining the Research Object.** The power source of RF system is taken as the research object. The indenture levels are as follows.

Initial indenture level: power module.

Minimum indenture level: constitutional unit:  
 $P = \{p_i\}$ ,  $i = 1, 2, \dots, 6$ , including input filter circuit, DC/DC converter 1, DC/DC converter 2, output filter circuit 1, output filter circuit 2, and output filter circuit 3.

**3.1.2. Dividing Product Levels.** The power source module is divided into levels according to its structure and functional characteristics, and the hierarchical division diagram is shown in Figure 7.

**3.1.3. Importance Analysis.** The power source converts 270 V DC to 28 V DC output and provides overheating protection for DC/DC and short-circuit protection for input filter. The power source adopts the redundant form of dual DC/DC converters output in parallel and isolated from each other. That is to say, when two DC/DC converters work normally, they will share current output; when one DC/DC converter is damaged, the other one will output full power (500 W), which reduces the impact on the normal operation of power module. At the same time, 28 V DC at the output end is divided into two separate output channels. Among them, the output end of output filter 2 is the normal output end of power supply, while output filter 3 only provides a short-term power transfer of about 30 seconds through short-term capacitor storage when the normal power source



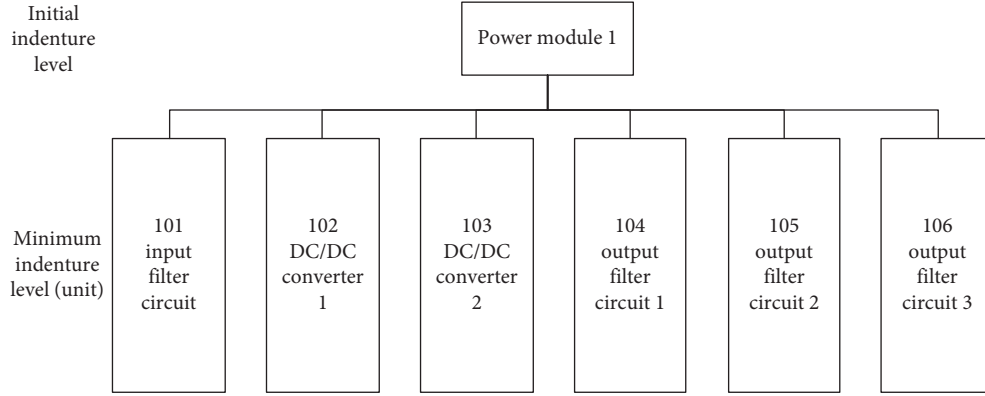


FIGURE 7: Schematic diagram of power module hierarchy.

terminal fails, which is not normally used. Its function diagram is shown in Figure 8.

The functional decomposition diagram of the power source product is shown in Figure 9.

According to the functional characteristics and engineering experience of step-down power source product, experts in the field scored each subfunction of the power module.

The difference in function importance of RF power source is not too big, and the number of subfunctions is not too much, so the 0–4 scoring method is used as the scoring standard [22]. The final scoring results are shown in Table 1.

Finally, the importance of each function of power module to the system is  $I_1 = 0.46$ ,  $I_2 = 0.29$ ,  $I_3 = 0.11$ , and  $I_4 = 0.14$ , as shown in Table 2.

The product function should be realized by the entity, and each subfunction is corresponding to its entity. As shown in the function decomposition in Figure 5, the function importance of the entity unit  $P = \{p_1, p_2, \dots, p_6\}$  is expressed as

$$\begin{aligned}
 I_f(p_1) &= I_2 + I_4 \\
 &= 0.43, \\
 I_f(p_2) &= I_1 + I_4 \\
 &= 0.6, \\
 I_f(p_3) &= I_1 + I_4 \\
 &= 0.6, \\
 I_f(p_4) &= I_2 \\
 &= 0.29, \\
 I_f(p_5) &= I_2 \\
 &= 0.29, \\
 I_f(p_6) &= I_3 \\
 &= 0.11.
 \end{aligned} \tag{10}$$

Therefore, the function importance weight of each unit power module is  $\delta_i(p_i) = I_f(p_i) / \sum_1^4 I_f(p_i)$ ,  $i = 1, 2, \dots, 6$ , i.e., input filter circuit, DC/DC conversion circuit 1, DC/DC conversion circuit 2, output filter circuit 1, output filter

circuit 2, and output filter circuit 3 are  $\delta_i(p_i) = \{0.19, 0.26, 0.26, 0.12, 0.12, 0.05\}$ .

According to the above importance analysis, all component objects  $P = \{p_k\}$ ,  $k = 1, 2, \dots, 6$  are sorted according to the importance of power module, and we get  $P^* = \{\text{input filter circuit, output filter circuit 2, DC/DC conversion circuit 1, DC/DC conversion circuit 2, output filter circuit 1, output filter circuit 3}\}$ . Output filter 3 only provides a short-term power transfer function for the power source when the normal output of the power source fails. Its state does not affect the normal operation of the power module, so the weight of the function importance is very small. Therefore, according to the requirements, we can focus on the key units of high importance for detailed characterization analysis.

**3.1.4. Characterization Factor Analysis.** In this paper, FMEA and failure rate statistical analysis are carried out for each key unit of power source module. Combined with engineering experience and expert opinions, the main failure modes, fault correlation, and monitoring parameters are determined. The simplified FMEA analysis table is shown in Figure 10.

Furthermore, according to the statistical failure rate distribution of historical fault case data, the frequency of DC/DC converter output failure accounts for the largest proportion of power system failure, and the failure modes of overheating protection, output short circuit, or output overvoltage caused by converter failure in DC/DC converter also occupy a large proportion. Therefore, power health characterization parameters mainly include DC/DC output voltage, ripple, insulation resistance, and module temperature.

Combined with domestic and foreign research, engineering experience, and expert opinions, six key functional units  $P^* = \{p_i\}$ ,  $i = 1, 2, \dots, 6$ . According to the importance analysis, the importance of output filter 3 to the system is very low, which basically does not affect the normal operation of the power supply. Therefore, without considering the transient transfer function, the set of key health characterization factors can be expressed as  $X_k = \{\text{power conversion efficiency } \eta, \text{ DC/DC1 temperature } T1, \text{ DC/DC2 temperature } T2, \text{ power output voltage } u, \text{ power output}$

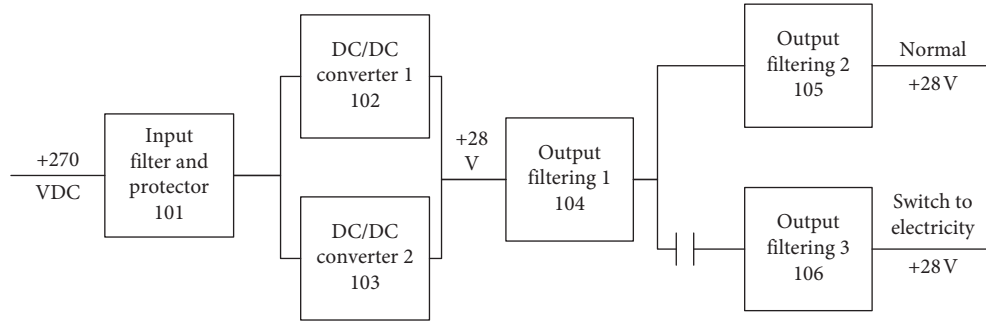


FIGURE 8: Functional block diagram of power source module.

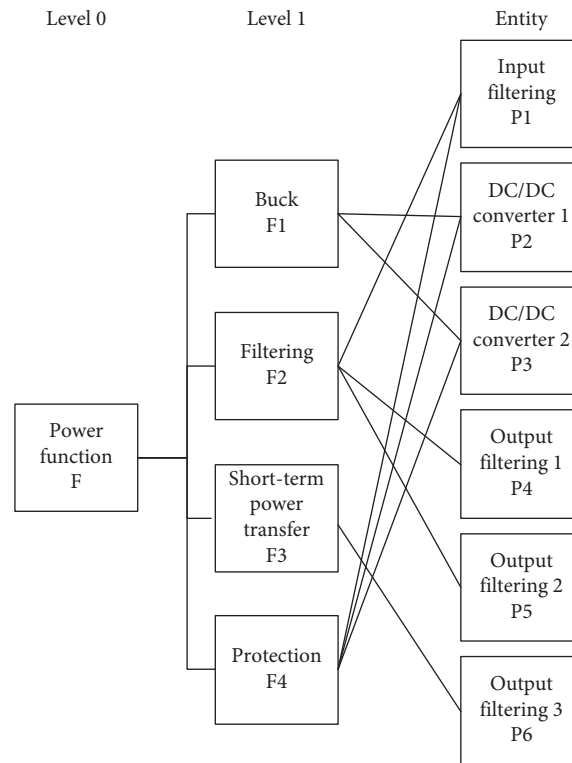


FIGURE 9: Functional decomposition diagram of power source module.

TABLE 1: 0–4 scoring standard table.

| Comparison score<br>$k_i^j$ | Importance                      |                       |                        |                            |                                |
|-----------------------------|---------------------------------|-----------------------|------------------------|----------------------------|--------------------------------|
|                             | A is much more important than B | A is important than B | A is as important as B | A is not as important as B | A is far less important than B |
| $A(k_A^B)$                  | 4                               | 3                     | 2                      | 1                          | 0                              |
| $B(k_B^A)$                  | 0                               | 1                     | 2                      | 3                          | 4                              |

voltage ripple  $V$ }, which can cover or transform the above four factors greatly.

#### 4. Health State Assessment of RF Power Supply

Product health includes functional integrity, temporal sustainability, and environmental adaptability. Among them, the integrity of functions is the foundation and main basis

for measuring the product health. Therefore, the health of RF power source is directly related to the completeness of its functions. In addition, it should have the stability to deal with all kinds of disturbances and be able to work reliably for a long time so as to achieve the health in a broad sense.

By using functional integrity, signal-to-noise ratio, and reliability, we establish a three-dimensional spatial model of product health in terms of function, environment, and time

TABLE 2: Functional importance based on functional decomposition.

|       | F1 | F2 | F3 | F4 | Accumulated score | Correction score | Functional importance $I_i$ |
|-------|----|----|----|----|-------------------|------------------|-----------------------------|
| F1    | —  | 4  | 4  | 4  | 12                | 13               | 0.46                        |
| F2    | 1  | —  | 3  | 3  | 7                 | 8                | 0.29                        |
| F3    | 0  | 0  | —  | 2  | 2                 | 3                | 0.11                        |
| F4    | 0  | 1  | 2  | —  | 3                 | 4                | 0.14                        |
| Total |    |    |    |    | 24                | 28               | 1.00                        |

Note. In order to avoid zero score for the least important function, one point can be added to the cumulative score of each function as the correction score, and then the next step of calculation can be carried out.

Initial indenture level:

Power module

Minimum indenture level: unit

| Code | Product or function unit name | Failure mode   | Failure cause   | Failure effect  |  | SEV | Class | Fault correlation or detection parameters     |
|------|-------------------------------|--|---|---|--|-----|-------|---|
|      |                               |  |   | Local effect  | Upper effect   |     |       |   |
| 101  | Input filter circuit          | 28V no output  | Fuse FU1 open circuit; Diode V2 open circuit; Capacitor C1 ~ C5, C10 ~ C13 short circuit; L1 open circuit | The input filter circuit has no output                                    | Power module 28V no output   | III | E     | Output voltage                                |
|      |                               | Output ripple overrun                                | Filter capacitor open circuit   | Input filter circuit output ripple overrun                                | Nothing  | IV  | E     | Output ripple                                 |
|      |                               | Short circuit between output and housing             | Capacitor C8, C9 short circuit  | The output of the input filter circuit is short circuited to the housing  | Power module electrical strength and insulation resistance index overrun | III | E     | Insulation resistance and electrical strength |
|      |                               | Input reverse connection protection function failure | Diode V2 open circuit   | Input filter circuit input reverse connection protection function failure | Power module input reverse connection protection function failure        | III | E     | Output voltage                                |

FIGURE 10: The screenshot of FMEA.

and comprehensively describe the health of RF power source from three aspects of integrity, stability, and reliability. First of all, a high temperature test was conducted on a certain type of RF power supply. The duration was 72 hours. A group of health characterization parameter data of power source was collected every 20 minutes. Each group contained 5 key characterization parameter data. A total of 200 groups of observation sequence samples were collected. Among them, the first 20 groups of observation data under the initial normal state are taken as the health reference state data, and the last 180 groups of data are taken as the observation data under different test conditions.

**4.1. Functional Integrity.** Suppose that the main function of the power source is  $F_i$ ,  $i \in R$ , including step-down, filtering, overheating, and short-circuit protection. To judge whether its function is in good condition, it can be obtained by detecting whether the voltage, ripple, and efficiency exceed the failure threshold. The functional integrity C is only divided into 0 and 1. However, when considering the short-term transfer function, it is also necessary to detect the voltage at the transfer terminal, so the basic characterization parameters of the power source are needed for the functional integrity. According to the previous function importance analysis, the function weights of step-down, filter, power

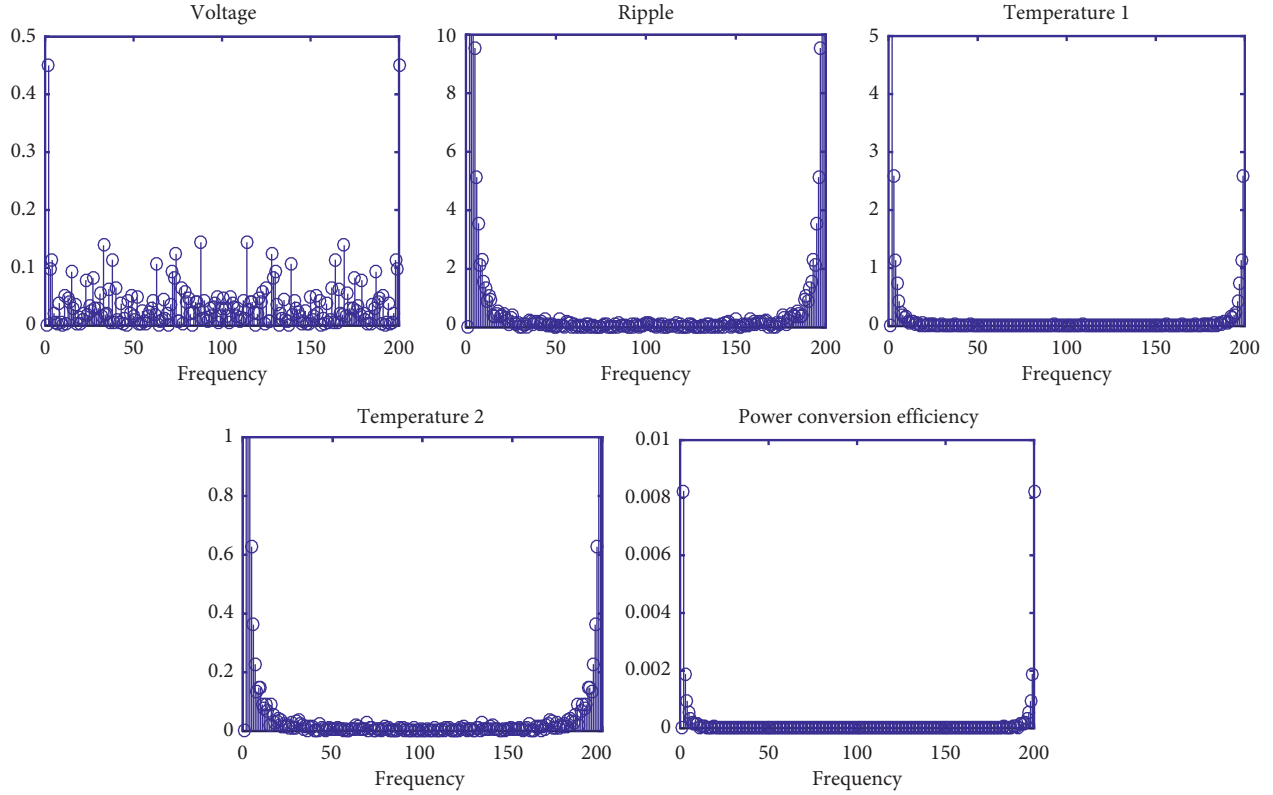


FIGURE 11: Power spectrum of power source key characteristic parameters based on fast Fourier transform.

transfer, and overheat protection are  $I_1 = 0.46$ ,  $I_2 = 0.29$ ,  $I_3 = 0.11$ , and  $I_4 = 0.14$ .

Therefore, the level threshold of functional integrity can be determined according to the importance of each function. When all functions can be executed normally, the functional integrity of the power source is 1; when the short-term power transfer function fails, the product function integrity can be determined by formula (2). By substituting the data into the above formula, it can be concluded that the power source functional integrity is 0.89. Obviously, the power

transfer failure does not affect the normal operation of the power supply, so 0.89 is taken as the threshold value of degradation stage; when the filtering function fails, the functional integrity decreases to 0.71, and the power source function is very incomplete, and the voltage output is unstable and close to the fault, which belongs to the upper limit of danger, and the step-down function fails. Therefore, it can be considered that when the functional integrity is 0.54, it has entered the area with the most incomplete function. The area is divided as follows:

$$\begin{aligned}
 C &= \sum_{i=1}^4 I_i \|F_i \\
 &= \begin{cases} \{0.89, 1\}, & \text{functionally intact,} \\ \{0.75, 0.86\}, & \text{functional decline,} \\ \{0.57, 0.60, 0.71\}, & \text{dangerous area,} \\ \{0, 0.11, 0.25, 0.40, 0.43, 0.46, 0.54\}, & \text{the lowest functional integrity.} \end{cases} \quad (11)
 \end{aligned}$$

According to the 200 sets of characteristic parameter data collected in 72 hours, it is found that none of them has actually reached the failure threshold (such as output voltage less than 25V, greater than 29V, and ripple peak  $\geq 120$  mV), and there is no functional failure. Therefore, from the perspective of whether the overall function of the power

source can be implemented, the functional integrity has always been  $\prod_{i=1}^4 \|F_i = 1$ .

4.2. *Environmental Adaptability.* Taking the key health characterization parameters of the product as the time-domain

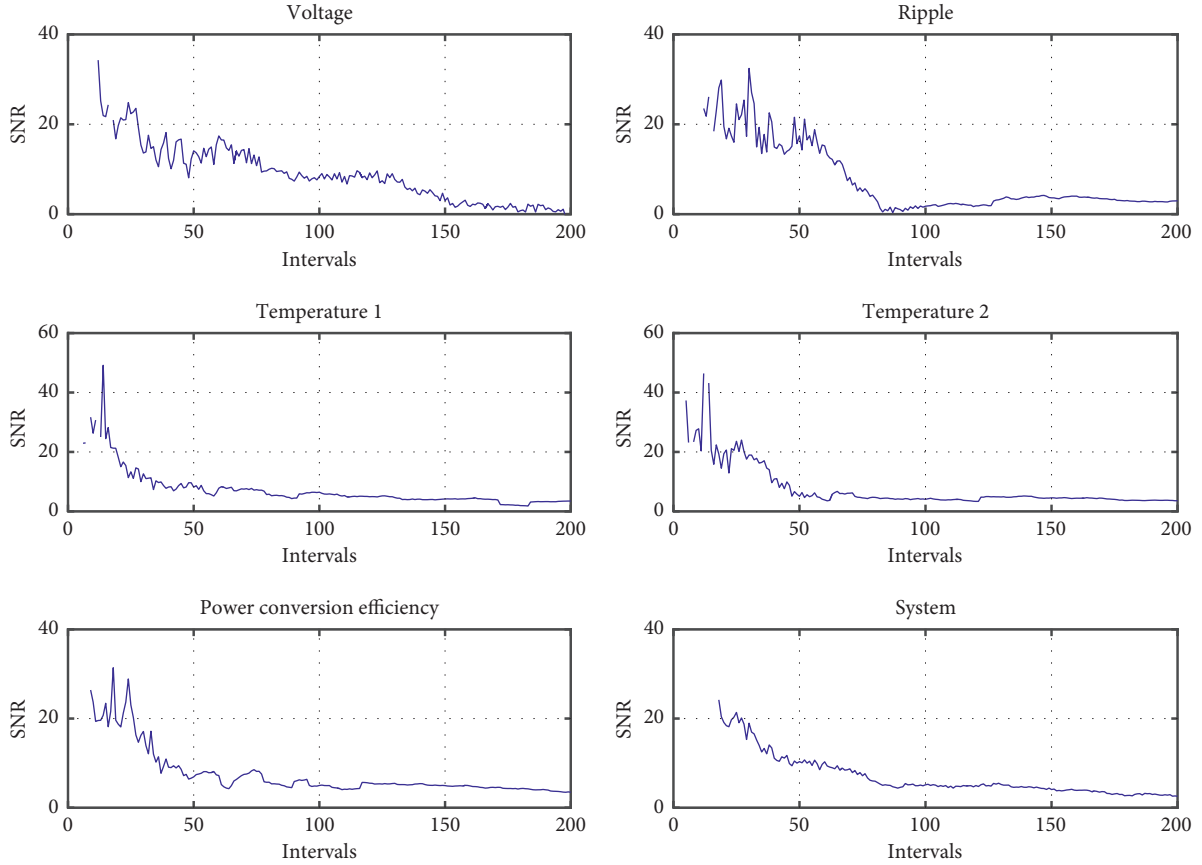


FIGURE 12: The power source key characteristic parameters and the system comprehensive signal-to-noise ratio curve.

signal, the health characterization parameters under the environment can be converted to the frequency domain through fast Fourier transform (FFT). The spectrum is shown in Figure 11.

According to formula (6), the SNR of each parameter is obtained, as shown in Figure 12.

The results show that when the power source is running at high temperature, the signal-to-noise ratio  $S_N$  shows a downward trend, indicating that the proportion of signal and noise of the product increases gradually, and the anti-interference and stability ability of the product to the environment gradually decreases, that is, the environmental adaptability decreases. It should be noted that from the beginning of sampling, the signal-to-noise ratio (SNR) presents a downward trend until it reaches the 40th sampling point. At this time, the signal-to-noise ratio (SNR) is about 14–16dB, and 15 dB is taken as the threshold value of SNR degradation stage; then, the downward trend gradually stabilizes until the sample points of 150–160th group decrease slowly, and the signal-to-noise ratio is about 3–6dB, so 5 dB is the signal-to-noise ratio with better environmental adaptability. According to the definition of SNR, when SNR is less than 0 dB, the signal is far less than noise, so 0 dB is taken as the worst threshold of environmental adaptability.

Thus, according to the signal-to-noise ratio, the level region of environmental adaptability can be determined. When the signal-to-noise ratio  $S_N > 15$ , the environmental adaptability is the best; when the signal-to-noise ratio

$5 < S_N \leq 15$ , the environmental adaptability decreases, but it is acceptable; when the signal-to-noise ratio  $0 < S_N \leq 5$ , the product has low environmental adaptability and is in the dangerous area; when the signal-to-noise ratio  $S_N \leq 0$ , the environmental adaptability of the product is the worst and extremely unstable.

**4.3. Temporal Sustainability.** Temperature is one of the important environmental factors affecting product reliability. Research [23] shows that the failure rate of electronic product increases monotonously with temperature according to exponential law when the temperature is higher than general indoor environment temperature (about 20°C–25°C). It is known that the MTBF of the power source module is 718.7041 hours when it operates at 100°C high temperature. Since the power source belongs to electronic product, its service life generally follows exponential distribution, so

$$\begin{aligned}
 \lambda(t) &= \lambda, \\
 \text{MTBF} &= \frac{1}{\lambda}, \\
 R(t) &= e^{-\lambda t} \\
 &= e^{-t/\text{MTBF}},
 \end{aligned} \tag{12}$$

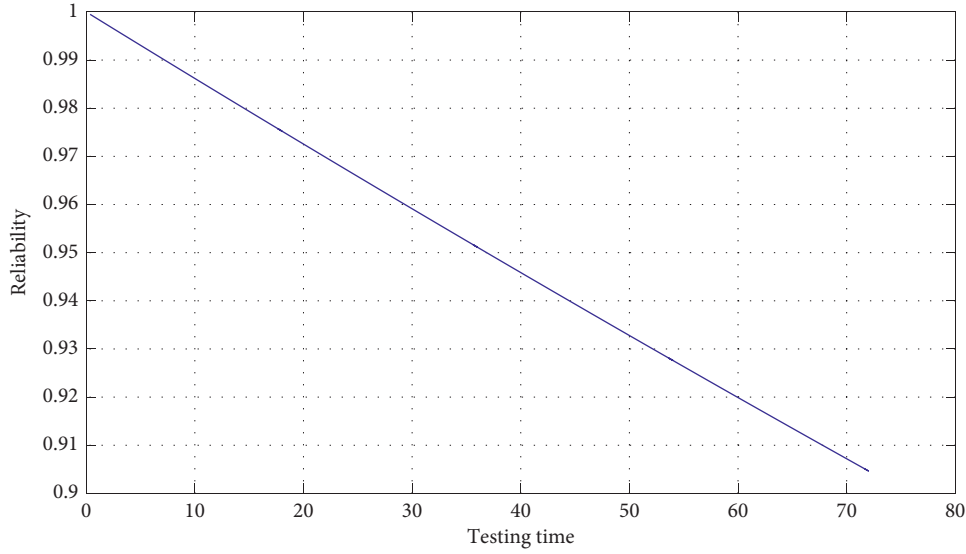


FIGURE 13: Reliability curve of aging test stage.

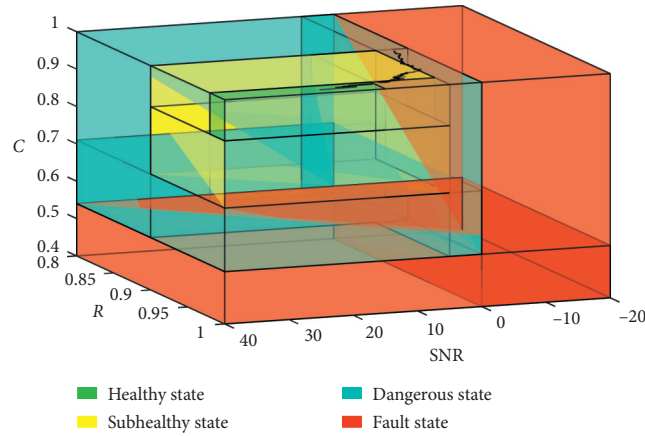


FIGURE 14: Health grade and health curve in three-dimensional space.

where  $R(t)$  is the reliability and  $\lambda(t)$  is the failure rate.

When there is a power supply, the failure rate can be approximated as

$$\lambda = \frac{1}{\text{MTBF}} = 1.3914 \times 10^{-3}, \quad (13)$$

and the power reliability function is obtained as follows:

$$R(t) = e^{-\lambda t} = e^{-1.3914 \times 10^{-3} t}. \quad (14)$$

According to the above reliability function, the reliability value of the corresponding time in 72-hour experiment is obtained, as shown in Figure 13.

With time increasing, the reliability decreases continuously, which indicates that the reliable continuous operation time is decreasing, so the product temporal sustainability

decreases. According to the engineering experience and the characteristics of electronic devices, the health threshold of reliability is 0.98, while for electrical devices, the product failure rate with reliability of 0.90 is quite high, so only the risk threshold is set as 0.90.

#### 4.4. Three-Dimensional Complex Measurement Model.

Through the analysis of the above three dimensions, the health level area can be divided according to the parameter thresholds (functional integrity, signal-to-noise ratio, and reliability)—with three dimensions of functional integrity, environmental adaptability, and temporal sustainability. Therefore, the health level  $l$  can be divided into healthy state, subhealthy state, dangerous state, and fault state.

$$l = \begin{cases} \{1 \geq C > 0.89, S_N > 15, 1 \geq R > 0.98\}, & \text{healthy state,} \\ \{0.71 < C \leq 0.89, S_N > 5, 0.9 < R \leq 1\} \cup \\ \{0.71 < C \leq 1, 5 < S_N \leq 15, 0.9 < R \leq 1\} \cup & \text{subhealthy state,} \\ \{0.71 < C \leq 1, S_N > 5, 0.9 < R \leq 0.98\}, \\ \{0.54 < C \leq 0.71, S_N \geq 0, 0 \leq R \leq 1\} \cup & \text{dangerous state,} \\ \{0.54 < C \leq 1, 5 \geq S_N \geq 0, 0 \leq R \leq 1\}, \\ \{0 \leq C < 1, S_N < 0, 0 \leq R \leq 1\} \cup & \text{fault state.} \\ \{0 \leq C < 0.54, -\infty < S_N < +\infty, 0 \leq R \leq 1\}, \end{cases} \quad (15)$$

The RF power supply health level partition of the three-dimensional space is shown in Figure 14. At this time, the health level is no longer a plane point line, but a number of three-dimensional areas. The green area represents the healthy state, yellow represents the subhealthy state, blue represents the dangerous state, and red represents the fault state. According to the values of functional integrity  $C$ , signal-to-noise ratio  $S_N$ , and reliability  $R$ , a healthy three-dimensional spatial expression  $H_\Delta = (C, S_N, R)$  can be obtained, as shown in the black curve in the figure, and then the health degree of the power source at each time can be determined according to the health level.

As can be seen from the above figure, with the power supply aging, the health curve of the three-dimensional space gradually changes from the initial green area to the yellow area and then to the blue area, which represents the gradual evolution of the health of the power source to subhealth and then to the dangerous state.

## 5. Conclusion

In this paper, a new three-dimensional complex measurement model-based avionic radio-frequency power source health assessment method is proposed. The functional integrity, signal-to-noise ratio, and reliability are taken as the parameter indexes of functional integrity, environmental adaptability, and temporal sustainability. Taking a certain RF power source as an example, the three-dimensional complex health assessment model which can clearly reflect the three dimensions of RF power source is then built, respectively. The proposed method also provides a new approach to solve the product health state uncertainty under multiple factor effects and optimization management for decision makers. However, further research can be carried out to find an accurate solution for threshold division in the process of health measurement.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## References

- [1] N. Fushun, "Some knowledge on fault prediction and health management technology," *Acta Instrumenta Sinica*, vol. 39, no. 8, pp. 1–14, 2018.
- [2] C. Deng, Y. Z. Sun, R. Li et al., "Hidden markov model based on the heavy-duty CNC health state estimate," *Jisuanji Jicheng Zhizao Xitong/Computer Integrated Manufacturing Systems, CIMS*, vol. 19, no. 3, pp. 552–558, 2013.
- [3] J. Wang, K. Fan, and W. Wang, "Integration of fuzzy AHP and FPP with TOPSIS methodology for aeroengine health assessment," *Expert Systems With Applications*, vol. 37, no. 12, 2010.
- [4] X. Ren, L. Chen, DeS. Li et al., "Fault decision of computer numerical control machine system using grey clustering analysis and rough set theory," vol. 11, no. 5, 2019.
- [5] K. Zhu and X. Song, "Cross-fuzzy entropy-based approach for performance degradation assessment of rolling element bearings," *Proceedings of the Institution of Mechanical Engineers*, vol. 232, no. 2, 2018.
- [6] D. Wang, H. Wang, Q. Wang, and H. Liu, "Comprehensive health assessment of marine diesel generator set based on analytic hierarchy process," *Electromechanical Equipment*, vol. 37, no. 3, pp. 53–58, 2020.
- [7] A. H. Tai, W.-Ki Ching, and L. Y. Chan, "Detection of machine failure: hidden markov model approach," *Computers & Industrial Engineering, Advances in Mechanical Engineering*, vol. 57, no. 2, 2008.
- [8] L. Bin, S. Dong, H. Lin et al., "Complex system health assessment based on Mahalanobis distance and bin-width estimation technique," *Chinese Journal of entific Instrument*, vol. 37, no. 9, 2016.
- [9] Y. Liao, Y. Xiang, and M. Wang, "Health assessment and prognostics based on higher order hidden semi-markov models," 2020, <http://arxiv.org/abs/2002.05272v1>.
- [10] B. Samanta, "Gear fault detection using artificial neural networks and support vector machines with genetic algorithms," *Mechanical Systems and Signal Processing*, vol. 18, no. 3, 2004.
- [11] C. Xu, H. Zhang, D. Peng, Y. Yu, C. Xu, and H. Zhang, "Study of fault diagnosis of integrate of D-S evidence theory based on neural network for turbine," *Energy Procedia*, vol. 16, 2012.
- [12] J. Liu and E. Zio, "System dynamic reliability assessment and failure prognostics," *Reliability Engineering and System Safety*, vol. 160, 2017.
- [13] F. Qin, J. Chen, X. Zhang, M. Jin, J. Zheng, and X. Li, "Research on GIS running state evaluation based on FKNN

- algorithm,” *Journal of Hubei University of Technology*, vol. 33, no. 2, pp. 62–66, 2018.
- [14] A. Oluwasegun and J.-C. Jung, “The application of machine learning for the Prognostics and Health Management of control element drive system,” *Nuclear Engineering and Technology*, vol. 52, 2020.
- [15] W. U. Ke-Xiong, W. Zhen-Hua, and S. Unit, “Health assessment and analysis method based on expert system and data driven,” *Heilongjiang Science*, vol. 10, no. 16, 2019.
- [16] X. Zhang, H. F. Sun, and J. X. ZhaoNi, “Bearing health assessment based on Hilbert transform envelope analysis and cluster analysis,” *International Journal of Reliability and Safety*, vol. 13, no. 3, p. 151, 2019.
- [17] B. Lu and Yi Hu, “Discussion on environmental adaptability evaluation technology of avionics,” *Science and Technology and Innovation*, vol. 6, no. 9, pp. 96-97, 2020.
- [18] H. Don, “Johnson. Signal-to-noise ratio,” *Scholarpedia*, vol. 1, p. 2088, 2006.
- [19] J. F. De Boer, B. Cense, B. H. Park, M. C. Pierce, G. J. Tearney, and B. E. Bouma, “Improved signal-to-noise ratio in spectral-domain compared with time-domain optical coherence tomography,” *Optics Letters*, vol. 28, pp. 2067–2069, 2003.
- [20] D. B. Hawkins and W. S. Yacullo, “Signal-to-Noise ratio advantage of binaural hearing aids and directional microphones under different levels of reverberation,” *Journal of Speech and Hearing Disorders*, vol. 49, pp. 278–286, 1984.
- [21] GJB451-90, *Terms of Reliability and Maintainability*, 1990.
- [22] W. Hongyan, “Application of 0-4 scoring method in technical and economic analysis of sewage treatment system,” *Science and Technology Innovation Guide*, vol. 14, no. 34, pp. 136–138, 2017.
- [23] X. Wang, Z. Ren, and Q. Wang, *Study On The Relationship Between Failure Mechanism And Environmental Stress Of Avionics*, pp. 120–124, Kunming, China, 2006.



## Research Article

# Multiple-Devices-Process Integrated Scheduling Algorithm with Time-Selective Strategy for Process Sequence

Xiaohuan Zhang <sup>1</sup>, Dan Zhang,<sup>1</sup> Zhen Wang,<sup>1</sup> and Yu Xin<sup>2</sup>

<sup>1</sup>College of Computer Science and Technology, Huizhou University, Huizhou 516007, China

<sup>2</sup>Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China

Correspondence should be addressed to Xiaohuan Zhang; zhangxiaohuan@hzu.edu.cn

Received 7 August 2020; Revised 6 September 2020; Accepted 3 October 2020; Published 17 October 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Xiaohuan Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper is in view of the current algorithm to solve Multiple-Devices-Process integrated scheduling problems without considering parallel process between the parallel processing. Meanwhile, the influence of the first processing procedure on the second processing procedure is ignored, resulting in poor tightness between serial processes and poor parallelism between parallel processes, ultimately affecting the product scheduling results. We proposed Multiple-Devices-Process integrated scheduling algorithm with time-selective for process sequence. The proposed Multiple-Devices-Process sequencing strategy determines the process of scheduling order and improves the tightness between serial process. This paper presents a method to determine the quasischeduling time point of the multiequipment process, the time-selective strategy of Multiple-Devices-Process, and the time-selective adjustment strategy of Multiple-Devices-Process so that the first and the second processing processes cooperate with each other, and the purpose of improving the tightness of the serial process and the parallelism of the parallel processes is achieved, so as to shorten the product processing time.

## 1. Introduction

Scheduling, as a key factor affecting the production efficiency of enterprises, has always been a hot issue studied by scholars. Efficient scheduling optimization algorithm can maximize production efficiency and help enterprises achieve higher benefits [1]. At present, there are two main research directions in this field, namely, single processing (assembly) scheduling and integrated scheduling. Typical representatives of the former scheduling are job-shop and flow-shop scheduling [2]. This type of processing and assembly is common in mass production. The reason is the large quantity of the ordered products. The centralized production of disassembling them into work pieces will generate a large amount of inventory, and then the inventory will be assembled. In this way, synchronous processing and assembly can shorten the production cycle.

Currently, there are many research achievements in this direction, and more advanced methods include genetic

algorithm [3], tabu search [4], neural network [5], heuristic algorithm [6], particle swarm optimization [7], bionics algorithm [8], and various hybrid algorithms [9]. With the improvement of people's living standards, today's consumers increasingly pursue personalized products. Personalized product orders are often single small batch orders, which only produce a small amount of inventory or even zero inventories in the production process.

As a result, the above production methods will no longer be dominant but will separate the internal relationship between processing activities and assembly activities. The integrated scheduling problem makes up for this defect and provides practical solutions for single small batch production [10]. The main scheduling idea is that during the scheduling process, the assembly can be carried out as long as the work piece is finished and the assembly conditions are satisfied. The assembly activity is carried out synchronously with the ongoing processing activity, so as to effectively improve the production efficiency and reduce internal

consumption [11]. The most advanced research achievements in this field include the integrated scheduling algorithm for multiple-devices-process.

The content of this paper is the scheduling problem of Multiple-Devices-Process in the integrated scheduling problem. In the actual manufacturing process, some processes need to be processed by multiple equipment. This paper will study this problem. At present, the research on this problem is in its infancy, puts forward this problem, and gives solutions for the first time. In this method, the critical path method and priority scheduling the Multiple-Devices-Process method are adopted to determine the process scheduling sequence, and the first adaptive strategy is applied to determine the scheduling time of the process. The disadvantages are as follows: (1) Paying too much attention to the serial processing of the process, not considering the parallel processing between parallel processes. (2) The influence of the first processing step on the second processing step is ignored, resulting in poor tightness between serial processes and poor parallelism between parallel processes, which ultimately affects the product scheduling result.

In order to solve the above problems, a Multiple-Devices-Process Integrated Scheduling Algorithm with Selecting Time for Process Sequence is proposed (MDOI-SAWSTFPS). The Multiple-Devices-Process sequencing strategy is proposed. It can determine the scheduling sequence of the process and improve the tightness between serial processes. This paper presents a method to determine the quasischeduling time point of the Multiple-Devices-Process, the time-selective strategy of Multiple-Devices-Process, and the time-selective adjustment strategy of Multiple-Devices-Process so that the first and the second processing processes cooperate with each other, and the purpose of improving the tightness of the serial process and the parallelism of the parallel processes is achieved, so as to shorten the product processing time.

## 2. Problem Description and Analysis

The product processing technology of the multidevices scheduling problem is shown as a tree structure. The node in the tree represents the working procedure of the product, the directed edge represents the partial order relation of the working procedure processing order, the root node represents the last working procedure in the product, and the finished processing of the root node indicates the finished processing of the product. The following constraints must be met for the scheduling problem of Multiple-Devices-Process:

- (1) Each process must strictly comply with the agreed partial order relationship in the processing tree.
- (2) Each equipment can only process one process at any time, and the processing process can not be interrupted.
- (3) There are no devices with the same function in the device set.

- (4) There are multiequipment processes in the process set in which there are several related types of equipment working together. The multidevices process refers to the process that requires multiple equipment to cooperate with each other in a process. The processing time of multiple equipment is the same, and the processing start time and the processing end time are the same.
- (5) A process can only be processed if and only if all the preceding processes are in the finished state (or no prior process).
- (6) The difference between the end time of the latest processing step and the beginning time of the earliest processing step is the total processing time of the product.

## 3. Strategy Design

*3.1. Multiple-Devices-Process Sequencing Strategy.* The design of this strategy is similar to that of the common equipment process sequence sequencing strategy. The difference between the two strategies is that the multidevices process sequencing strategy contains multidevices processes and needs to be discussed separately. Since the start time, end time, and processing time of these processes are the same, they can be regarded as the same process, and they are arranged together in no particular order [12]. For example, the product process tree is shown in Figure 1, which contains 7 processes. First, the path length of 3 leaf nodes is calculated, and the results are A2:35, A6:70, and A7:80. Therefore, the process sequence 1 is A1, A3, A5, and A7. Remove the process in the first process sequence from the process tree, and the remaining leaf node is A2, A6. The path length is calculated, respectively, and the results are A2:20 and A6:35. Therefore, the second process sequence is A4, A6. Remove the process in the second process sequence from the process tree, and the remaining leaf node is A2. Since it is the last process sequence, there is no need to calculate its path length, and the upper process of the third process sequence is directly determined as A2.

A simplified version of the proposed algorithm can be described as follows:

Step 1:  $i = 0$ .

Step 2: calculate the path length of existing leaf nodes in the reverse process tree, respectively.

Step 3:  $i++$ .

Step 4: select the leaf node  $W$  with the longest path.

Step 5: if process  $W$  is not unique, select the process  $Q$  with the highest number of sequences on its path.

Step 6: if process  $Q$  is not unique, select the process  $O$  with the smallest difference between the number of layers of the process and the number of layers of the root node in the original processing tree.

Step 7: if process  $O$  is not unique, select the process  $P$  with the maximum total processing time of all preceding processes of each process on its path.

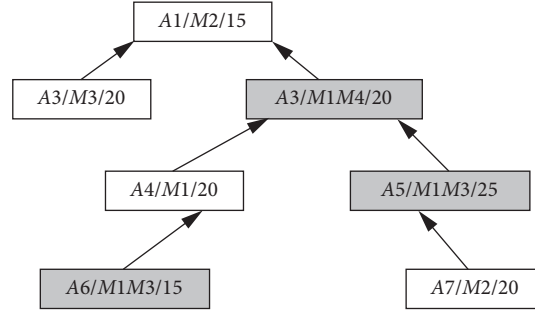


FIGURE 1: A product process tree with Multiple-Devices-Process.

Step 8: the sequence formed by all the processes on the path of process  $P$  is denoted as process sequence  $I$ .

Step 9: starting from process  $P$ , all the processes on process sequence  $I$  are successively pushed into  $S$ , followed by bomb stack  $S$ , and the resulting processes are successively stored in queue  $Qu$ .

Step 10: remove the process in queue  $Qu$  from the process tree.

Step 11: judge whether the current process tree is empty or not. Turn step 2 and turn step 12.

Step 12: exit.

**3.2. Determination of Quasischeduling Time Points of Multiple-Devices-Process.** Set the current scheduling process as process  $A$ , the finishing time of its pretightening process is  $T$ . Starting from point  $T$  on its processing equipment, the processing end time point of each scheduled process is found as the “quasischeduling time point” of process  $A$  and added to the “quasischeduling time point” set.

The determination of quasischeduling time point of Multiple-Devices-Process needs to consider the processing equipment of each parallel subprocess separately. The set of quasischeduled time points of a Multiple-Devices-Process is the union of all parallel subprocesses on its processing equipment. For example, there are two parallel subprocesses in processes  $A$ ,  $A1$ , and  $A2$ , which are processed on different processing equipment, respectively. The “quasischeduling time point” set of  $A1$  is  $Ta1 = \{T1, T2, T3\}$ . The “quasischeduling time points” set of  $A2$  is  $Ta2 = \{T4, T5\}$ . The set of “quasischeduling time points” of procedure  $A$  is  $tA = ta1 \cup ta2$ .

**3.3. Time-Selective Strategy of Multiple-Devices-Process.** The time-selective strategy of Multiple-Devices-Process schedules each parallel subprocess separately at each time point in the process’s “quasischeduled time points” set. In the scheduling process, it should be noted that the first quasischeduling time point is the end time of the pretight process in the processing process tree of this process. There are two situations at this time point, as illustrated below.

As shown in Figure 2,  $Wi$  is a Multiple-Devices-Process, which contains two parallel subprocesses  $Wi1$  and  $Wi2$ .

According to the partial order relationship in the process tree, the start time of process  $Wi1$  and  $Wi2$  should be consistent and greater than or equal to the end time  $T1$  of the process which tightening front of process  $A$  in the process tree. The processing equipment of process  $Wi1$  is  $M2$ , and the processing equipment of process  $Wi2$  is  $M1$ . At this time, there are the following two conditions on the processing equipment of each parallel subprocess  $Wi$  at time  $T1$ .

**Situation 1:**  $T1$  in Figure 2(a) is the end time of the scheduled process  $B$  or blank time. At this point,  $T1$  is taken as the first “quasischeduling time point” of process  $Wi$ , so that  $Wi1$  starts processing on  $M2$  at time  $T1$ , and  $Wi2$  starts processing on  $M1$  at time  $T1$ .

**Situation 2:**  $T1$  in Figure 2(b) is the processing time of the scheduled procedures  $B$  and  $C$ . In order to avoid conflict, there are two solutions. One is processes  $B$  and  $C$  do not move. Select  $T3$ , the maximum end time of steps  $B$  and  $C$ , as the first “quasischeduling time point” of step  $Wi$ . Second,  $T1$  is selected as the first “quasischeduling time point” of process  $Wi$ , and process  $B$  is adjusted to the back of process  $Wi1$ , and process  $C$  is adjusted to the back of process  $Wi2$ . As shown in Figure 3, it is obvious that  $Wi$  is scheduled according to the second processing mode, and the resulting Multiple-Devices-Process trial scheduling scheme set will contain the case of  $Wi$  scheduled according to the first processing mode. To enlarge the problem solution space, the second processing method is selected to schedule  $Wi$ .

A simplified version of the proposed time-selective strategy of the Multiple-Devices-Process algorithm can be described as follows:

Step 1: queue a process from queue  $Qu$  and set to process  $A$ .

Step 2: suppose the end time of the tight preprocess of procedure  $A$  is  $T$ .

Step 3: determine if the process is an ordinary process or a Multiple-Devices-Process. If the ordinary process, go to 4. If the Multiple-Devices-Process, go to 5.

Step 4: start from point  $T$  on the processing equipment of process  $A$ , find the finishing time point of each

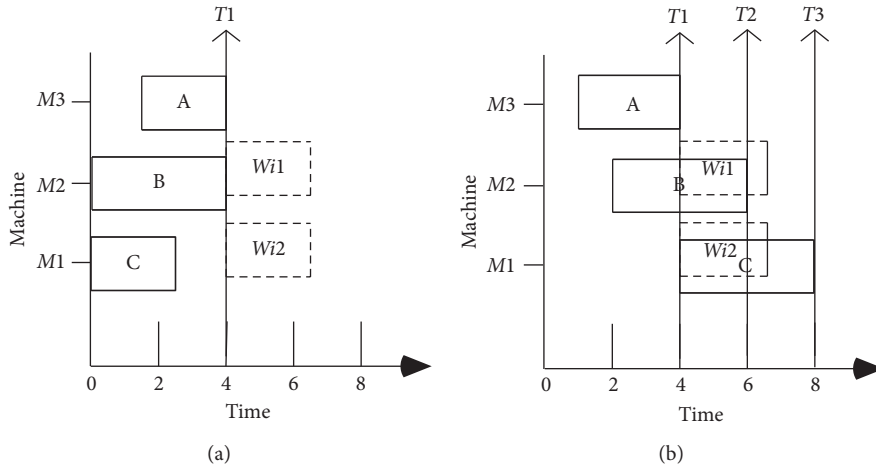


FIGURE 2: Scheduling analysis at the first quasischeduling time point.

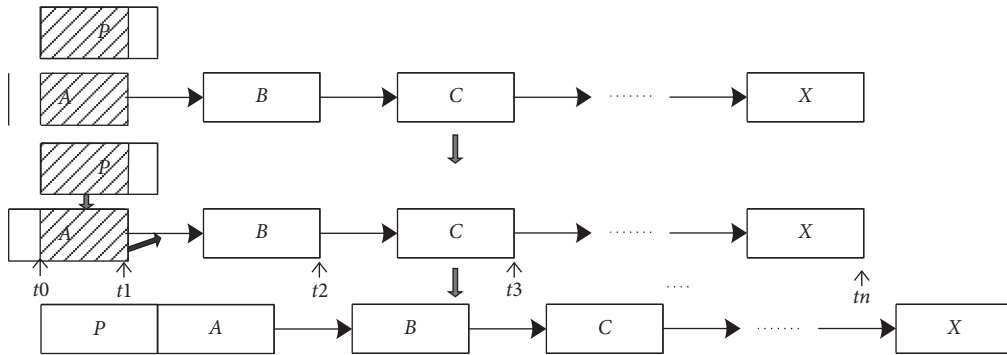


FIGURE 3: A schematic diagram of the backshift of the process affected by scheduling at the first “quasischeduling time point.”

scheduled process as the “quasischeduling time point” of process *A*, and add it into the “quasischeduling time point” set.

Step 5: starting from point *T* on several parallel processing equipment of process *A*, the processing end time point of each scheduled process on each parallel processing equipment is found to be the “quasischeduling time point” of process *A* and added to the “quasischeduling time point” set.

Step 6: judge whether the set of “quasischeduling time points” is empty, not for idling to 7, for idling to 15.

Step 7: take a time point *T* from the set of “quasischeduled time points,” and judge whether *T* point is the first “quasischeduled time point,” if it goes to 8, if not to 11.

Step 8: if process *A* is an ordinary process, go to 9. If process *A* is a Multiple-Devices-Process, go to 10.

Step 9: determine whether process *A* has affected the scheduled process on the current equipment. If the affected process is added to the list of scheduled processes of the equipment, the position is after process *A* and goes to 12.

Step 10: judge whether each parallel subprocess of Multiple-Devices-Process *A* has affected the scheduled process on the current device. If the affected procedure is added to the scheduled process chain list of each parallel subprocess equipment, the position is after process *A* and then goes to 13.

Step 11: if process *A* is an ordinary process, go to 12. If process *A* is a multiple-devices-process, go to 13.

Step 12: take *T* as the start time of process *A* to conduct trial scheduling of process *A* and adjust the processes affected by the scheduling of process *A*, so as to produce the trial scheduling scheme generated by the scheduling of process *A* at the “quasischeduling time point,” and add the trial scheduling scheme into the set of trial scheduling schemes of process *A* and turn it to 10.

Step 13: take *T* as the start time of each parallel subprocess in process *A*, and then conduct trial scheduling for each subprocess, and adjust the processes affected by each parallel subprocess of the scheduling process *A* (in accordance with section 3.4). The trial scheduling scheme of process *A* at the quasischeduling time point is generated; the trial scheduling scheme is added to the trial scheduling scheme set of Process *A*.

Step 14: delete this time point in the quasischeduling time point collection and go to 6.

Step 15: exit.

**3.4. Time-Selective Adjustment Strategy of Multiple-Devices-Process.** The selection of “quasischeduling time point” only considers the processing order between processes, without considering whether the scheduling of the process will affect the scheduled process on the equipment. Therefore, after the scheduling process, it is necessary to check the scheduled process that may be affected by it. When the processing completion time of the scheduling process is greater than the processing start time of the subsequent process on the same equipment or the processing start time of the process in the process tree, the time-selective adjustment strategy shall be started. For Multiple-Devices-Process, all parallel subprocesses need to be checked separately. Considering the situation of the first quasischeduling time point, as shown in Figure 4(a), when scheduling process  $W_{i1}$ , the processes affected are: Process  $B$ , Process  $D$ , Process  $E$ , and Process  $F$ . Wherein, when process  $B$  determines the first quasischeduling time point, its equipment is processing process, which has been processed in the design of time-selective scheduling strategy. The method is to arrange it in the chain list of processed processes of the equipment, and the position is after process  $W_{i1}$ . To adjust process  $B$  will affect the work process  $D$  on the device after tight; to adjust process  $D$  will affect the work process  $E$  on the device after tight, and work process  $F$  after tight in the process tree, by the same token, the scheduling process  $W_{i1}$  affected by the process are process  $C$ , working process,  $G/I$  and  $H$ . Consider the case that it is not the first quasischeduling time point, as shown in Figure 4(b), when scheduling process  $W_{i1}$ ; the processes affected are Process  $D$ , Process  $E$ , and Process  $F$ . When scheduling process  $W_{i1}$ , the processes affected are Process  $C$ , Process  $I$ , and Process  $H$ . To sum up, the time-selective adjustment strategy considers two situations when adjusting a process: one is the posttightening process on the processing equipment of the scheduling process. The other is the posttightening process on the equipment of the adjusted process and the middle-tightening process in the processing technology tree.

A simplified version of the proposed Multiple-Devices-Process timing adjustment strategy algorithm can be described as follows:

Step 1: queue all parallel subprocesses of the current Multiple-Devices-Process in sequence into  $Aqu$ .

Step 2: queue  $Aqu$  and store the results in  $W$ .

Step 3: judge whether  $W$  is empty, do not idle to 4, otherwise go to 10.

Step 4: suppose there are  $k$  post-tight processes in  $w$ 's process tree,  $n = 1$ .

Step 5: judge whether  $n > k$  is true. If it is true, go to 8. If not, go to 6.

Step 6: judge whether the processing completion time of  $W$  is greater than the processing start time of process  $wn$  after the tight process in the process tree (the starting time of the unscheduled process is  $+\infty$  by default), if to 7, otherwise to 8.

Step 7: take the machining end time of  $w$  as the start time of  $wn$ , and put  $wn$  into the team  $Aqu$ ,  $n++$ .

Step 8: determine whether  $w$  has tight postprocess  $WM$  on the chain list of scheduled processes of its processing equipment and the completion time of  $W$  is greater than the start time of  $WM$ . If it is 9, otherwise it is 2.

Step 9: take the end time of  $W$  as the start time of  $WM$ , enter  $WM$  into  $Aqu$ , and turn it to 2.

Step 10: after the adjustment, calculate the total processing time of the current scheme, exit.

A simplified version of the proposed multiple-devices-process integrated scheduling algorithm with time-selective for process sequence can be described as follows:

Step 1: invert the processing partial order relationship in the process tree.

Step 2: determine the process scheduling sequence in the process tree by using the Multiple- Devices-Process sequencing strategy.

Step 3: process sequence 1 is queued out of queue  $Qu$  and an initial scheduling scheme is formed.

Step 4: determine whether the process queue  $Qu$  is empty, not for idling to 5, for idling to 10.

Step 5: queue a process from process queue  $Qu$  and set to process  $A$ .

Step 6: judge that process  $A$  is an ordinary process and multiple-devices-process. If it is an ordinary process, switch to 7. If it is a multidevices process, switch to 8.

Step 7: apply time-selective strategy and the time-selective adjustment strategy to schedule process  $A$ , generate process  $A$  trial scheme set, and go to 9.

Step 8: apply the time-selective strategy of Multiple-Devices-Process and the time-selective adjustment strategy of Multiple-Devices-Process to schedule process  $A$ , generate process  $A$  trial scheme set.

Step 9: the scheduling scheme with the minimum processing time is found out from the trial scheduling scheme set of process  $A$ , which is taken as the scheduling scheme of process  $A$ , go to 4.

Step 10: process  $A$  scheduling scheme is the product scheduling scheme.

Step 11: exit.

## 4. Example Analysis

In order to facilitate to the reader to understand the algorithm, the following is an example analysis. Product  $A$  is shown in Figure 5, and its reverse process tree is shown in

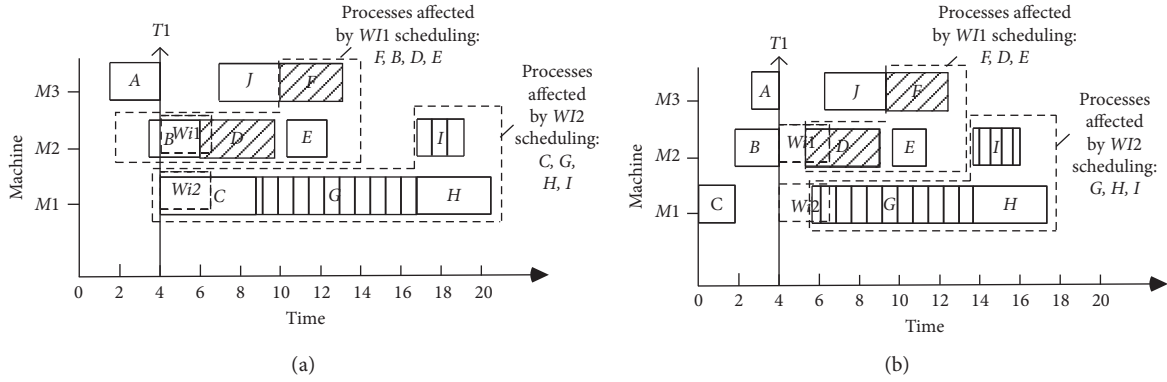


FIGURE 4: Scheduling analysis at the first quasischeduling time point.

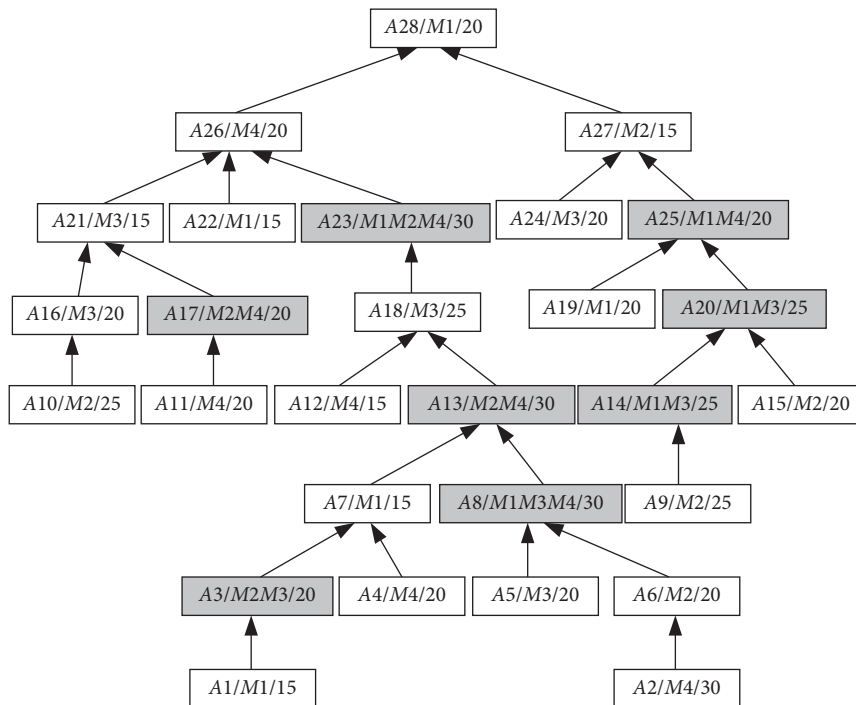


FIGURE 5: Product processing tree.

Figure 6 where each processing node represents  $A$  processing process, which is divided into ordinary processes and Multiple-Devices-Process. Such as  $A28/M1/20$ , which means that the process name is  $A28$ , and it is processed on equipment  $M1$ , which is an ordinary process and the processing time is 20.  $A17/M2M4/20$  means that the process name is  $A17$ , which needs to be processed on the equipment  $M2$  and  $M4$  at the same time, and it is a Multiple-Devices-Process. According to the algorithm proposed in this paper, the product processing process tree is shown in Figure 6. The scheduling process is shown in Table 1, and the total processing time of product  $A$  is 265. The result of scheduling product  $A$  is shown in Figures 7 and 8. Meanwhile, the total processing time of scheduling product  $A$  according to the algorithm proposed in [12] is 275, and the result of scheduling final product  $A$  is shown in Figure 9.

## 5. Experimental Results and Analysis

To verify the effectiveness of the proposed algorithm, four sets of data were randomly generated, with 50 products in each set. The parameters of the products were randomly generated. The software used on the experimental platform is Windows 10, 64 bit, GCC5.5, and the hardware of the experimental platform is an Intel Core I7-860 processor with 32 GB of memory. Five groups of experiments were designed as follows: Figure 10 is a comparison of the total elapsed processing time when the number of multidevices is 2 for the two algorithms, Figure 11 is a comparison of the total elapsed processing time when the number of multidevices is 3 for the two algorithms, Figure 12 is a comparison of the total elapsed processing time when the number of multidevices is 5 for the two

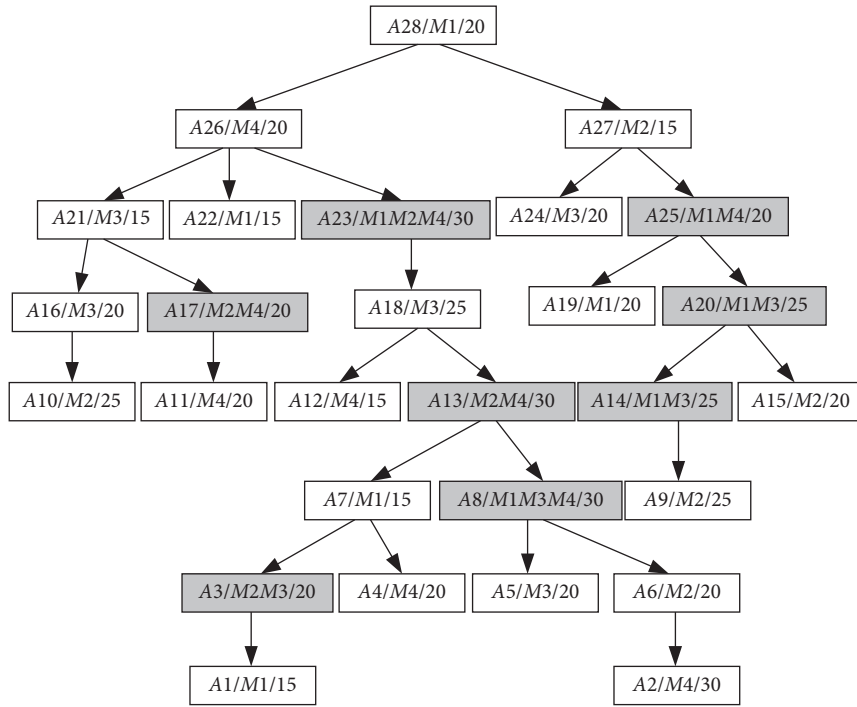


FIGURE 6: The reverse product processing tree.

TABLE 1: The scheduling process of the algorithm scheduling product A in this paper.

| Id  | Device | Quasischeduling time points | Total processing time of trial scheduling scheme | Determine the scheduling time point | Scheduling time points for each process in the current scheme   |
|-----|--------|-----------------------------|--|-------------------------------------|---|
| A28 | M1     | ---                         | ---  | 0                                   | A28:0   |
| A26 | M4     | ---                         | ---  | 20                                  | A28:0, A26:20   |
| A23 | M124   | ---                         | ---  | 40                                  | A28:0, A26:20, A23:40   |
| A18 | M3     | ---                         | ---  | 70                                  | A28:0, A26:20, A23:40, A18:70   |
| A13 | M4     | ---                         | ---  | 95                                  | A28:0, A26:20, A23:40, A18:70, A13:95   |
| A8  | M134   | ---                         | ---  | 125                                 | A28:0, A26:20, A23:40, A18:70, A13:95, A8:125   |
| A6  | M2     | ---                         | ---  | 155                                 | A28:0, A26:20, A23:40, A18:70, A13:95, A8:125, A6:155   |
| A2  | M4     | ---                         | ---  | 175                                 | A28:0, A26:20, A23:40, A18:70, A13:95, A8:125, A6:155, A2:175   |
| A27 | M2     | 20 70 125 175               | 205 205 205 205                                  | 20                                  | A28:0, A26:20, A23:40, A18:70, A13:95, A8:125, A6:155, A2:175, A27:20   |
| A25 | M1M4   | 35 70 95 125 155 205        | 220 205 220 220 205 225                          | 70                                  | A28:0, A26:20, A23:40, A18:70, A13:95, A8:125, A6:155, A2:175, A27:20, A25:70   |
| A20 | M1M3   | 90 95 155                   | 250 205 205                                      | 95                                  | A28:0, A26:20, A23:40, A18:70, A13:95, A8:125, A6:155, A2:175, A27:20, A25:70, A20:95                                   |
| A14 | M1M3   | 120 155                     | 225 205  | 155                                 | A28:0, A26:20, A23:40, A18:70, A13:95, A8:125, A6:155, A2:175, A27:20, A25:70, A20:95, A14:155                          |
| A9  | M2     | 180                         | 205  | 180                                 | A28:0, A26:20, A23:40, A18:70, A13:95, A8:125, A6:155, A2:175, A27:20, A25:70, A20:95, A14:155, A9:180                  |
| A21 | M3     | 40 95 120 155 180           | 205 215 215 220 205                              | 40                                  | A28:0, A26:20, A23:40, A18:70, A13:95, A8:125, A6:155, A2:175, A27:20, A25:70, A20:95, A14:155, A9:180, A21:40          |
| A16 | M3     | 55 95 120 155 180           | 210 220 220 225 205                              | 180                                 | A28:0, A26:20, A23:40, A18:70, A13:95, A8:125, A6:155, A2:175, A27:20, A25:70, A20:95, A14:155, A9:180, A21:40, A16:180 |

TABLE 1: Continued.

| Id  | Device | Quasischeduling time points      | Total processing time of trial scheduling scheme | Determine the scheduling time point | Scheduling time points for each process in the current scheme  |
|-----|--------|----------------------------------|--|-------------------------------------|--|
| A10 | M2     | 200 205                          | 250 230  | 205                                 | A28:0, A26:20, A23:40, A18:70, A13:95, A8:125, A6:155, A2:175, A27:20, A25:70, A20:95, A14:155, A9:180, A21:40, A16:180, A10:205   |
| A7  | M1     | 125 155 180                      | 245 245 230                                      | 180                                 | A28:0, A26:20, A23:40, A18:70, A13:95, A8:125, A6:155, A2:175, A27:20, A25:70, A20:95, A14:155, A9:180, A21:40, A16:180, A10:205, A7:180   |
| A3  | M2M3   | 205 230                          | 250 250  | 205                                 | A28:0, A26:20, A23:40, A18:70, A13:95, A8:125, A6:155, A2:175, A27:20, A25:70, A20:95, A14:155, A9:180, A21:40, A16:180, A10:205, A7:180, A3:205   |
| A1  | M1     | 225                              | 250  | 225                                 | A28:0, A26:20, A23:40, A18:70, A13:95, A8:125, A6:155, A2:175, A27:20, A25:70, A20:95, A14:155, A9:180, A21:40, A16:180, A10:205, A7:180, A3:205, A1:225   |
| A17 | M2M4   | 55 70 90 125 155 175 205         | 285 270 265 270 270 265 270                      | 90                                  | A28:0, A26:20, A23:40, A18:70, A13:110, A8:140, A6:170, A2:190, A27:20, A25:70, A20:95, A14:170, A9:195, A21:40, A16:195, A10:240, A7:195, A3:220, A1:240, A17:90  |
| A11 | M4     | 110 140 170 220                  | 285 285 265 265                                  | 170                                 | A28:0, A26:20, A23:40, A18:70, A13:110, A8:140, A6:170, A2:190, A27:20, A25:70, A20:95, A14:170, A9:195, A21:40, A16:195, A10:240, A7:195, A3:220, A1:240, A17:90, A11:170   |
| A4  | M4     | 210 220                          | 265 265  | 210                                 | A28:0, A26:20, A23:40, A18:70, A13:110, A8:140, A6:170, A2:230, A27:20, A25:70, A20:95, A14:170, A9:195, A21:40, A16:195, A10:240, A7:195, A3:220, A1:240, A17:90, A11:170, A4:210   |
| A24 | M3     | 35 55 95 120 170 195 215 240     | 265 265 265 265 285 280 280 265                  | 35                                  | A28:0, A26:20, A23:40, A18:70, A13:110, A8:140, A6:170, A2:230, A27:20, A25:70, A20:95, A14:170, A9:195, A21:55, A16:195, A10:240, A7:195, A3:220, A1:240, A17:90, A11:170, A4:210, A24:35   |
| A19 | M1     | 90 120 170 195 210 255           | 265 265 285 275 265 275                          | 90                                  | A28:0, A26:20, A23:40, A18:70, A13:110, A8:140, A6:170, A2:230, A27:20, A25:70, A20:110, A14:170, A9:195, A21:55, A16:195, A10:240, A7:195, A3:220, A1:240, A17:90, A11:170, A4:210, A24:35, A19:90                                    |
| A15 | M2     | 135 140 190 220 240 265          | 310 265 280 285 285 285                          | 140                                 | A28:0, A26:20, A23:40, A18:70, A13:110, A8:140, A6:170, A2:230, A27:20, A25:70, A20:110, A14:170, A9:195, A21:55, A16:195, A10:240, A7:195, A3:220, A1:240, A17:90, A11:170, A4:210, A24:35, A19:90, A15:140                           |
| A5  | M3     | 170 195 215 240                  | 285 280 280 265                                  | 240                                 | A28:0, A26:20, A23:40, A18:70, A13:110, A8:140, A6:170, A2:230, A27:20, A25:70, A20:110, A14:170, A9:195, A21:55, A16:195, A10:240, A7:195, A3:220, A1:240, A17:90, A11:170, A4:210, A24:35, A19:90, A15:140, A5:240                   |
| A22 | M1     | 40 70 90 110 135 170 195 210 255 | 280 280 275 275 275 280 270 265 270              | 210                                 | A28:0, A26:20, A23:40, A18:70, A13:110, A8:140, A6:170, A2:230, A27:20, A25:70, A20:110, A14:170, A9:195, A21:55, A16:195, A10:240, A7:195, A3:220, A1:240, A17:90, A11:170, A4:210, A24:35, A19:90, A15:140, A5:240, A22:210          |
| A12 | M4     | 95 110 140 170 190 230 260       | 285 280 280 265 265 275 275                      | 170                                 | A28:0, A26:20, A23:40, A18:70, A13:110, A8:140, A6:170, A2:230, A27:20, A25:70, A20:110, A14:170, A9:195, A21:55, A16:195, A10:240, A7:195, A3:220, A1:240, A17:90, A11:185, A4:210, A24:35, A19:90, A15:140, A5:240, A22:210, A12:170 |



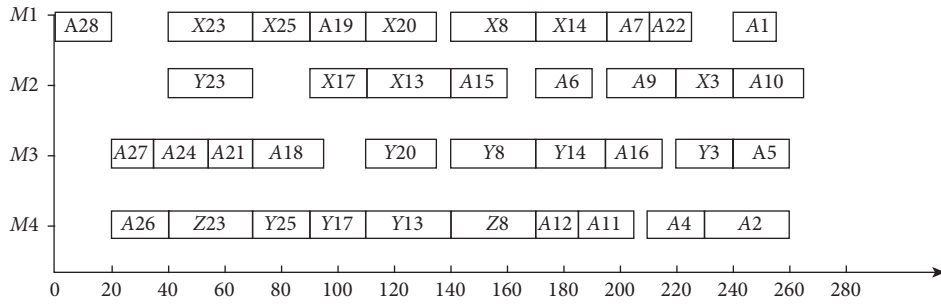


FIGURE 7: Scheduling results of the time-selective strategy of Multiple-Devices-Process.

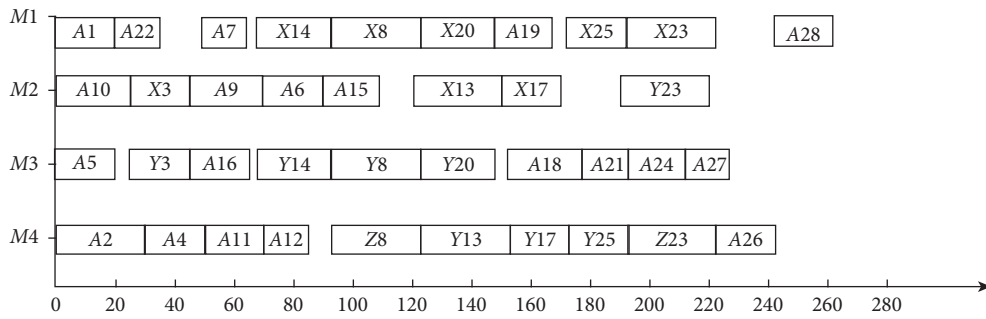


FIGURE 8: Reverse scheduling results of the time-selective strategy of Multiple-Devices-Process.

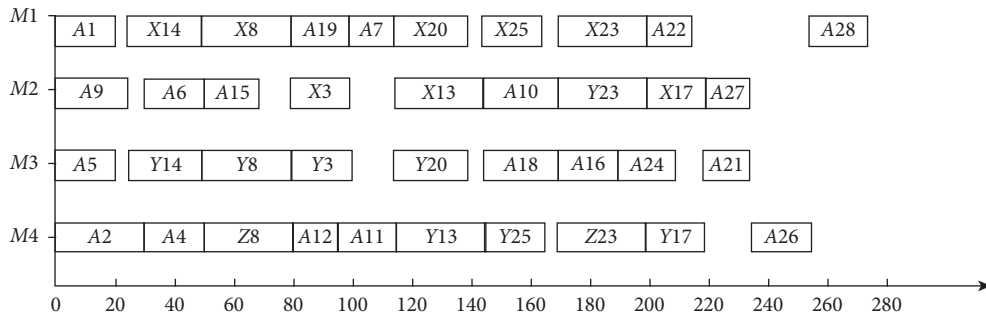


FIGURE 9: Scheduling results of the reverse multidevice first adaptation algorithm.

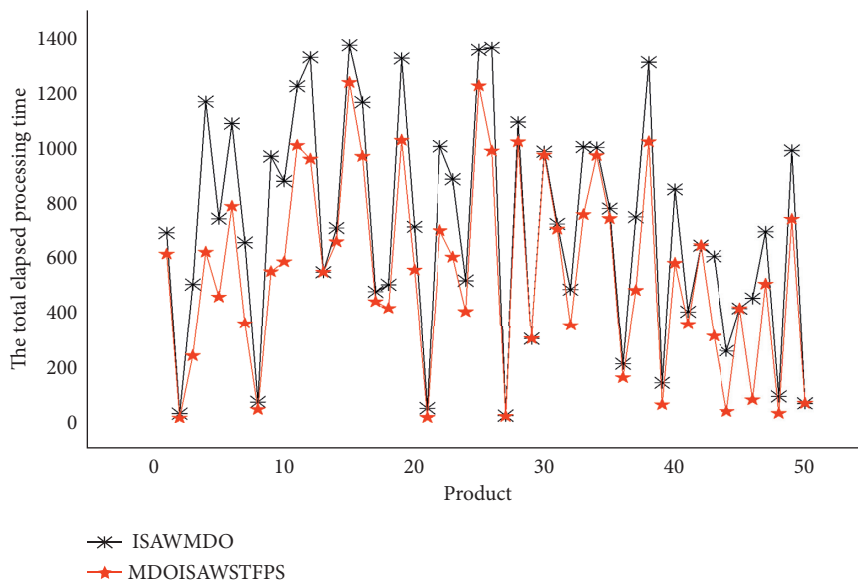


FIGURE 10: Comparison of the total elapsed processing time when the number of multiequipments is 2 for the two algorithms.

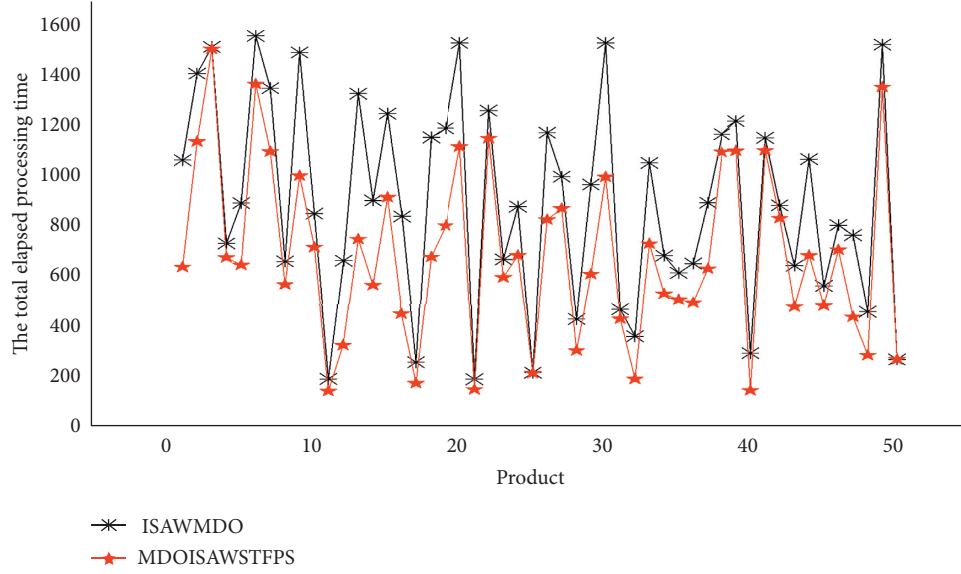


FIGURE 11: Comparison of the total elapsed processing time when the number of multiequipments is 3 for the two algorithms.

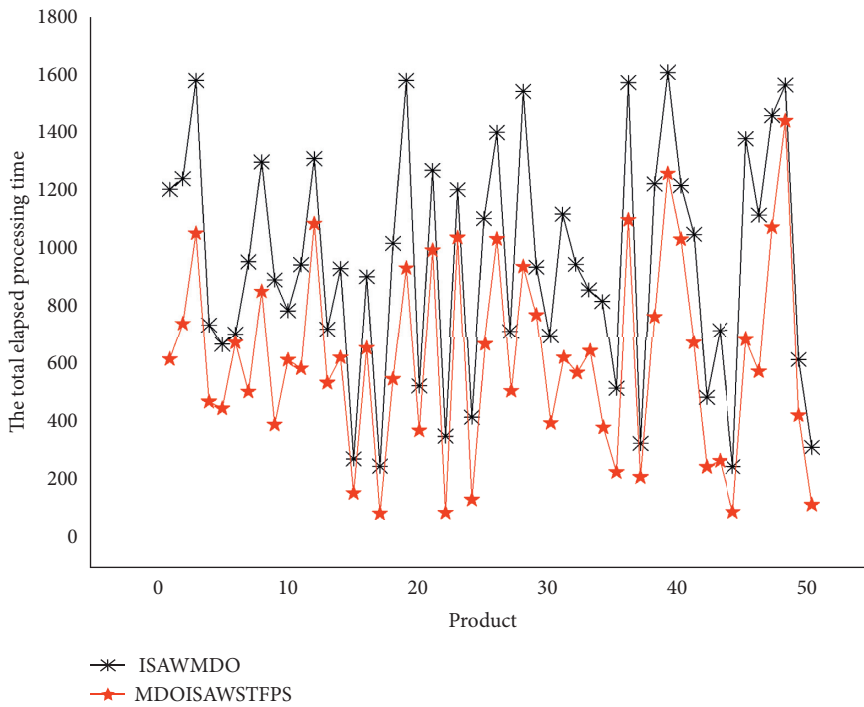


FIGURE 12: Comparison of the total elapsed processing time when the number of multiequipments is 5 for the two algorithms.

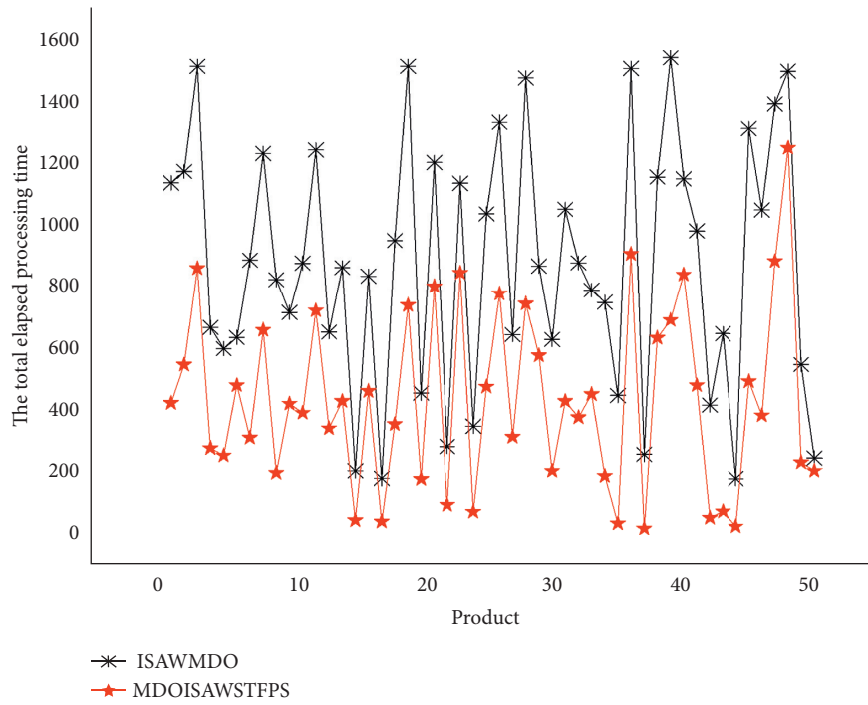


FIGURE 13: Comparison of the total elapsed processing time when the number of multiequipments is 10 for the two algorithms.

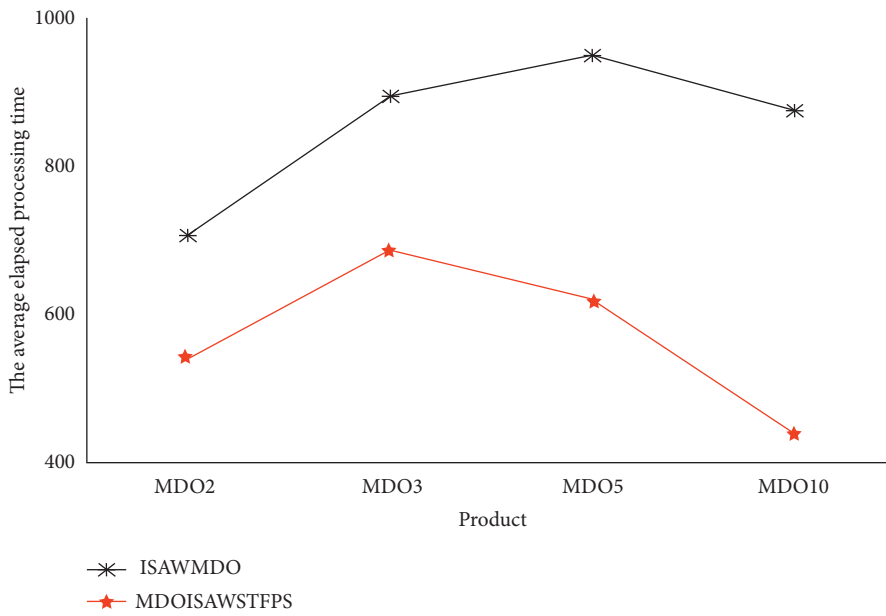


FIGURE 14: Comparison of the average elapsed processing time for the two algorithms.

algorithms, Figure 13 is a comparison of the total elapsed processing time when the number of multidevices is 10 for the two algorithms, and Figure 14 is a comparison of the average elapsed processing time for the two algorithms.

## 6. Conclusion

On the basis of ensuring parallel processing, the Multiple-Devices-Process sequencing strategy improves the tightness between serial processes and shortens the product completion time. The time-selective strategy of Multiple-Devices-Process and the time-selective adjustment strategy of Multiple-Devices-Process to make the current scheduling process has to cooperate with each other, to make the current partial product processing always take the smallest scheduling scheme, further shortening the product completion time, and when the scheme of minimum total available not only chooses the plan of process over time are the first, to further improve the possibility of parallel processing in the sequence after work. The scheduling results are superior to the existing Multiple-Devices-Process integrated scheduling algorithms. The introduction of a backtracking strategy to improve the accuracy of the algorithm may be the next direction of work, and it is intended to solve the problem of the multidevices process in distributed manufacturing.

## Data Availability

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also form part of an ongoing study.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

## Acknowledgments

This work was supported by the Project of Educational Commission of Guang dong (No. 2019KTSCX177), PhD Research Startup Foundation of Huizhou University (No. 2019JB014), Project of Ningbo Natural Science Foundation (No. 2019A610093), and the special research funding from the Marine Biotechnology and Marine Engineering Discipline Group in Ningbo University (No. 422004582).

## References

- [1] Q. Yuan, H. Tang, W. You, X. Wand, and Y. Zhao, "Virtual network function scheduling via multilayer encoding genetic algorithm with distributed bandwidth allocation," *Science China (Information Sciences)*, vol. 61, no. 9, pp. 93–111, 2018.
- [2] Y. Zhang, F. He, G. Lu, and H. Xioing, "An impossibility message scheduling based on modified genetic algorithm for time-triggered Ethernet," *Science China (Information Sciences)*, vol. 61, no. 1, pp. 279–281, 2018.
- [3] B. Zhou and X. Li, "Heuristic tabu search scheduling algorithm for wet-etching systems in semiconductor wafer fabrications," *High Technology Letters*, vol. 19, no. 2, pp. 111–116, 2013.
- [4] G. Mejía and J. Pereira, "Multiobjective scheduling algorithm for flexible manufacturing systems with Petri nets," *Journal of Manufacturing Systems*, vol. 1, no. 1, pp. 1–10, 2020.
- [5] X. Zhang and Y. Wang, "New mixed broadcast scheduling approach using neural networks and graph coloring in wireless sensor network," *Journal of Systems Engineering and Electronics*, vol. 20, no. 1, pp. 185–191, 2009.
- [6] X. Zhang, X. Cui, and S. Zheng, "Heuristic task scheduling algorithm based on rational ant colony optimization," *Chinese Journal of Electronics*, vol. 23, no. 2, pp. 311–314, 2014.
- [7] H. Ding and X. Gu, "Improved particle swarm optimization algorithm based novel encoding and decoding schemes for flexible job shop scheduling problem," *Computers & Process Research*, vol. 121, no. 1, 5 pages, Article ID 104951, 2020.
- [8] K. Kumar and E. Das, "Genetic algorithm and particle swarm optimization in minimizing make span time in job shop scheduling," *Proceedings of ICDMC-2019*, vol. 1, no. 1, 5 pages, 2020.
- [9] Y. . Jiao, Y. E. Chunming, and S. . Business, "Novel firefly algorithm for solving job shop scheduling problem," *Computer Engineering and Applications*, vol. 49, no. 11, pp. 213–203, 2013.
- [10] H. Zhang, L. Yang, and B. Liu, "Registration of multi-resolution medical images using a modified artificial fish-swarm algorithm combined with powell's method," *Journal of Xi'an Jiaotong University*, vol. 45, no. 4, pp. 46–52, 2011.
- [11] F. Naifar, M. Gzara, and M. T. Loukil, "A new leader guided optimization for the flexible job shop problem," *Journal of Combinatorial Optimization*, vol. 39, no. 10, pp. 1–5, 2020.
- [12] Z. Xie, Y. Qi, and J. Yang, "Integrated scheduling algorithm with multiple-devices- process," *Journal of Mechanical Engineering*, vol. 49, no. 6, pp. 89–98, 2014.

## Research Article

# Panoramic Display and Planning Simulation of Civil Engineering Project Based on Virtual Reality Technology

Zheng Wang <sup>1</sup>, Chang Liu,<sup>1</sup> and Bin Song<sup>2</sup>

<sup>1</sup>Academy of Fine Arts, Anyang Normal University, Anyang, Henan 455000, China

<sup>2</sup>E-Publishing Division, Tsinghua University Press, Anyang, Henan 455000, China

Correspondence should be addressed to Zheng Wang; 01278@aynu.edu.cn

Received 15 May 2020; Revised 27 September 2020; Accepted 5 October 2020; Published 17 October 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Zheng Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the traditional civil engineering industry, only abstract theoretical concepts are applied to express complex realistic contents, but this method of information transmission is sometimes very limited for engineering participants, because its expression is not comprehensive and specific and also makes the recipients have difficulty in understanding its information. The rise and development of virtual reality (VR) technology has become a good assistant to overcome this difficulty, providing an extremely effective means for displaying and simulating civil engineering project in real time. This paper introduced the methods and principles of VR platform system and structure dynamic analysis, conducted virtual scene design and three-dimensional modeling of civil engineering project, analyzed the synthetic debugging and simulation roaming of the virtual environment, proposed the engineering model of virtual construction process and planning structure calculation, discussed the realization method of virtual project management and visualization, and finally performed the simulation and its result analysis of panoramic display and planning of civil engineering project based on VR technology. The results show that the VR technology can simulate various environments and activities of civil engineering projects in advance, analyze the rationality of various schemes, and modify their different functional parameters, so it can achieve the simulation optimization of project planning, design, and construction.

## 1. Introduction

Virtual reality (VR) technology is an immersive interactive environment based on computable information, and it specifically uses modern high-tech with computer technology as the core to generate realistic visual, audio, and tactile integration in a specific range of virtual environments, through which the user interacts with the objects in the virtual environment in a natural way with the necessary equipment, so as to produce the feeling and experience of being in the real environment and realize the exchange of information between virtual and reality [1]. The VR is a more ideal form of human-computer interface between users and computers than traditional computer technology and users roam in a virtual environment and allow objects to be manipulated [2]. Applying the VR technology to the design, engineering control, and structural analysis of very large and complex structures will enhance the front-to-rear processing

capabilities of computing software. Compared with traditional computer technology, the VR technology has three important characteristics of immersion, interactivity, and imagination, and it has an extremely wide range of uses [3]. Because the VR technology can simulate various environments and activities beforehand, realize the function of analyzing the rationality of various schemes, and modify parameters of different schemes, it is possible to optimize planning, design, and construction [4, 5].

In the past, in the civil engineering industry, researchers could only express very rich content with very abstract theoretical concepts, such as graphic design drawings, section design drawings, elevation design drawings, and other flat design drawings to form some specific symbols, showing the three-dimensional building of three-dimensional images, and using abstract graphics and refined language as an auxiliary to describe the building to be built to convey information [6]. However, this method of

transmission will be restricted for the recipient of the information, because each recipient's work, level of knowledge, and ability to understand are different, so it will be difficult to communicate, and it will not be able to communicate what was originally intended [7]. The VR mobile terminals are applied to collect field data to achieve collaborative management of civil construction safety and quality and transfer potential safety hazards and quality defects on the site to the VR application in the form of data. The rise and development of VR technology has become a good assistant to overcome this difficulty, providing an extremely effective means [8]. This technology is a new high-end human-machine interface for the participants, and it is a three-dimensional environment generated by electronic computers, which not only enables users to feel the realistic existence of objects in the virtual world environment. As a structural entity, civil engineering is closely connected with the natural world and human environment, so the VR technology can be used to simulate the virtual environment [9].

Based on the summary and analysis of previous literature works, this paper expounded the research status and significance of VR technology, elaborated the development background, current situation, and future challenges of the panoramic display and planning simulation of civil engineering project, introduced the methods and principles of VR platform system and structure dynamic analysis, conducted virtual scene design and three-dimensional modeling of civil engineering project, analyzed the synthetic debugging and simulation roaming of the virtual environment, proposed the engineering model of virtual construction process and planning structure calculation, discussed the realization method of virtual project management and visualization, and finally performed the simulation and its result analysis of panoramic display and planning of civil engineering project based on VR technology. The detailed chapters are arranged as follows: Section 2 introduces the methods and principles of the VR platform system and structure dynamic analysis; Section 3 analyzes the panoramic display of civil engineering projects based on VR technology including the synthesis debugging and simulation roaming of virtual environment; Section 4 carries out the planning and simulation of civil engineering projects based on VR technology; Section 5 carries out civil engineering simulation and its result analysis; and Section 6 is the conclusion.

## 2. Methods and Principles

**2.1. VR Platform System.** The VR technology is based on the physiological and psychological characteristics of human vision and hearing and the computer generates realistic three-dimensional stereoscopic images. Users can put themselves in a virtual environment and become a member in the virtual environment by wearing interactive devices such as helmet displays and data gloves. The interaction between the user and various objects in the virtual environment is just like in the real world with all realistic feelings and there is a feeling of being on the scene. The human-

computer interaction in the VR system is a kind of nearly natural interaction. The user can not only use the computer keyboard and mouse to interact but also interact with sensor equipment such as special helmets and data gloves [10]. Users can investigate or operate objects in the virtual environment through their own natural skills such as language, body movement, or movement. Because the VR system is equipped with sensing and reaction devices for sight, hearing, and touch, users can obtain a variety of perceptions such as sight, hearing, and touch in the virtual environment, thereby achieving personal presence of feeling the environment. Due to the abovementioned advantages, the VR technology has been widely used in civil engineering and has broad application prospects.

The function realization of the simulation system is mainly at the model layer, the presentation layer realizes the interaction with the user, and the data layer realizes the recording and management of the simulation experiment data. According to the simulation teaching design, the main functions of user interaction are reflected in the movement of model components, rotation, and change of viewpoint and experiments of related animation display functions. Virtual reality modeling language is an object-oriented modeling language used to describe the data format of interactive objects and is also the standard of the VR technology. The modeling language can support network communication through the programming of its scripts, so that modeling language nodes can receive data through the network and can render scenes in real time. The language defines users' own objects by using prototypes, has object-oriented class characteristics, encapsulates corresponding data structures and methods, and is instantiated in applications, which has a certain degree of expansion. Through programming in routing and scripting languages, a visual scene can be constructed with a user interface, which not only facilitates user operations but also improves work efficiency. In the virtual simulation system, various common prototypes can be defined to realize the main simulation interactive functions, such as click, pan, rotate, select, and other functional modules.

Traditional geographic information systems have closed the characteristics from architecture to data format and different geographic information systems have different data storage formats. There are difficulties in data exchange between systems developed on different software, and data conversion standards can only partially solve the problem. Different application departments have different understandings of geographic phenomena and different data definitions of geographic information, which hinders the sharing of data between application systems. The digital city data have an infinitely seamless distributed data layer structure, including multisource, multiscale, multi-resolution, historical and current, vector format, and raster format data. In order to use these data for spatial analysis and decision-making, a spatial data warehouse must be established. The information sharing, data acquisition, and updating mechanisms and technologies have not been resolved, many necessary data standards and norms have not been established, and low-level repeated development has

occurred from time to time, resulting in a huge waste of human and material resources. Therefore, in order to achieve spatial information sharing, it depends on a special kind of metadata and data describing some attributes such as the format and precision of spatial data. The construction of the meta-database is very necessary; it is about the data of the database, through which you can understand the name, location, attributes and other information of the data, thereby greatly reducing the time for users to find data (Figure 1).

**2.2. Structural Dynamic Analysis Method.** The simulation model is composed of model components and the relationships between the components. The components are represented by multimedia objects, and the objects correspond to the entities in the real world; the interaction relationship between the components is realized by message events and the hierarchical relationship between them is described by combined multimedia simulation objects. The object is not only the basic conceptual unit of system analysis, but also the basic model unit of system design or basic programming unit of simulation programs. The object encapsulates all attribute definitions and operations about the object, and the state change of the object is triggered by internal or external events. Multimedia characteristics of sound that characterize the appearance of an entity, like quantitative parameters and state data that characterize the characteristics of an entity, are an integral part of real-world entity attributes. In real life, the external multimedia features of entities are an effective way for people to understand and understand nature; when people use models to describe the properties and state behavior of actual systems, they should not make the rich, colorful multimedia features of real objects. It is the most direct and simple way to lose sight of and disappear from the objective, so as to lose and understand the objective objects. The advanced modeling ideas and modeling methods of multimedia simulation are to ensure that this idea becomes a reality.

In the structural dynamic analysis method, the interface is firstly built and the first person and third person then are browsed through to allow users to get an immersive experience, getting a realistic effect brought by the VR technology and also getting a better understanding of every detail of the architectural plan. The flight mode is a bird-eye view of the underground buildings in the sky, which solves the problem that the overall style of the building cannot be observed in the building group and can judge whether it meets the aesthetic requirements and whether it can integrate with the surrounding buildings. Camera animation roaming is a roaming method that does not require manual operation and realizes browsing of multiple routes. This solves the problem of being unfamiliar with the overall architectural plan and easily disoriented and at the same time allows the user to relax. Finally, in the sound processing, the sound can deepen the sense of immersion and substitution, so that the entire operation process becomes more vivid, so that users get more operating experience. According to the above-mentioned functions to be realized, make buttons

corresponding to the functions, respectively. For the sound switch button, it is necessary to make an open button and a close button at the same position, monitor and replace each other, and execute the corresponding event [11].

The civil engineering works basically have the characteristics of large engineering volume and complex content, which is particularly noticeable in their construction plans. The framework and structure of virtual reality platform system is shown in Figure 2. The construction plan of the civil engineering project should take into account equipment management, material scheduling, capital operation, work type coordination, project schedule, quality control, cost control, information management, contract management, etc. Under such circumstances, the organization and management of construction plans are very difficult. Not only must they be able to collect, integrate, and use a large amount of engineering data, but also to balance contradictions and conflicts in various aspects. Therefore, the traditional organization and management of construction plans cost a lot of time and is difficult to ensure the quality and implement ability of the program. In civil engineering, a considerable part has the characteristics of a vast construction site and almost all civil engineering has three-dimensional features. In the traditional construction site organization and management, the floor plan is used to characterize and manage the construction site. Not only is it not intuitive but it also has to prepare a large number of plane drawings to fully display the three-dimensional characteristics of the site. The energy efficiency of organization and management is quite limited, and the organization and management of the construction site is a long-term and dynamic process, because during the construction process, the situation of the construction site will inevitably change with the progress of the project. The traditional site management uses paper drawings, which are difficult to modify and cannot reflect the dynamic changes of the site layout, and it is difficult to achieve the dynamic management requirements of the site.

### **3. Panoramic Display of Civil Engineering Projects Based on VR Technology**

**3.1. Virtual Scene Design and Three-Dimensional Modeling.** Panoramic technology is to stitch one or more groups of photos taken by the camera ring 360° into a panoramic image through seamless connection. The user applies special playback software to display it on the computer screen and can control the user's perspective through the mouse to achieve vertical or 360° horizontal scene viewing and supports arbitrary zoom-in and zoom-out, so that users can look around, look down, and look up as if they were in the field. After the design plan is determined, they perform virtual operations on the computer, and at this time, it can achieve the same situation as the actual construction and operation. After finding the problem, it can be quickly fed back to the planning and design to avoid the actual construction and operation process. If deficiencies and deficiencies are found, remedial measures will be taken, and some errors will cause great economic losses and the price of restricted use of

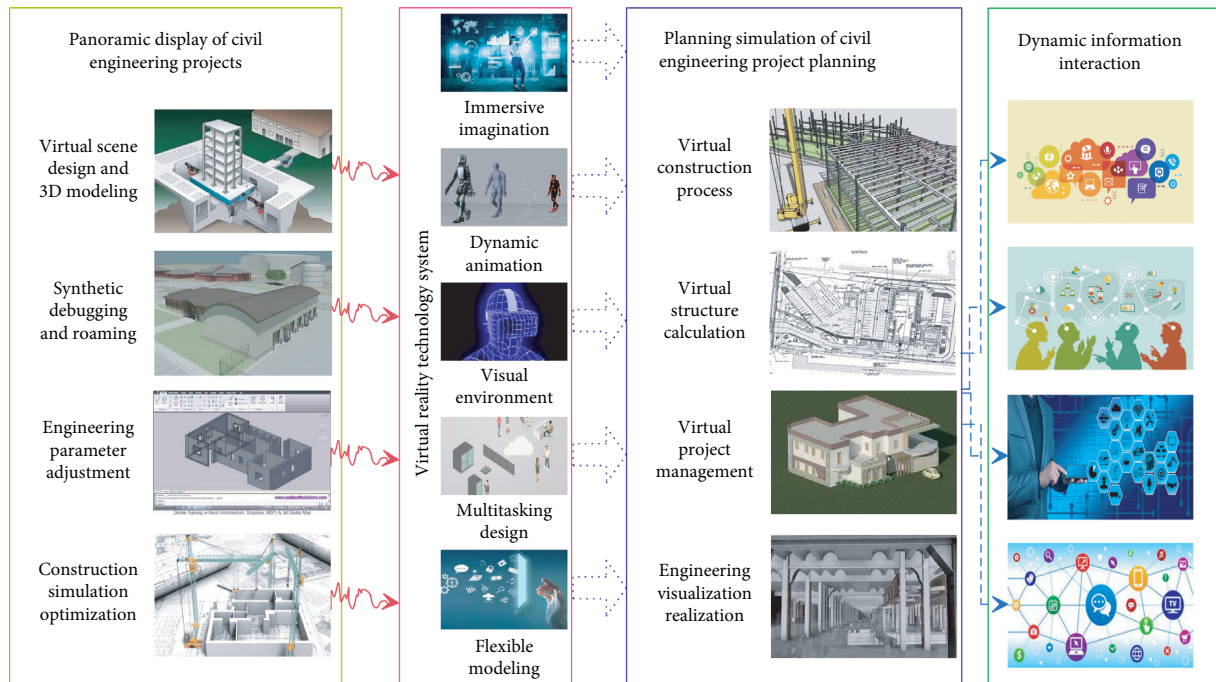


FIGURE 1: Virtual scene design and three-dimensional modeling of civil engineering project based on virtual reality technology.

functions. In the process of urban civil engineering design and implementation, the selection and optimization of construction schemes have a great impact on the construction quality and construction efficiency in the later stage. The effective application of the VR technology and the effective establishment of three-dimensional simulation models can to a certain extent effectively clarify the construction flow, construction procedures, etc., further optimize the selection of construction methods and construction machinery, and rationally arrange construction procedures.

In the construction process of civil engineering, the VR technology is mainly used for on-site construction management, to directly feedback the quality and safety issues on-site to construction management personnel, so as to be dealt with in time, improve construction quality, and to reduce the incidence of safety accidents (Figure 3). In practical applications, first log in to the VR application interface, then transfer the relevant photos of the construction site to the VR application, and in the corresponding options, including the project name, axis, and other specific parameters [12]. The VR mobile terminals are applied to collect field data to achieve collaborative management of civil construction safety and quality, transfer potential safety hazards and quality defects on the site to the VR application in the form of data, and then correlate these data to the VR model and the function of VR visualization to study on-site problems. In the construction of civil engineering construction, the use of VR technology can greatly improve the scientific and standardization of on-site construction and simplify the construction tasks. In addition, the use of BIM technology has also realized the supervision of the construction process, enhanced the use of information

technology, and reduced the interference of objective factors, which is conducive to comprehensive control of the project.

In the figure, different colored shapes represent different civil engineering construction entities; three red lines represent different perspective direction lines; different letters and numbers represent different virtual reality calibration points.

Traditional system simulation techniques rarely study the simulation of human perception models, and therefore cannot simulate human perception of the external environment. The introduction of the VR technology into the various stages of system simulation and simulation can make users immersed in it and have a clear understanding of the problems to be solved, instead of being limited to observing the simulation results from the screen, so that the establishment of the model and verification is more convenient. The VR technology is mainly reflected in the computer based on the established domain knowledge base and database and the use of artificial intelligence, pattern recognition, and other technologies, the main control agency for modeling, learning, planning, and calculation. Visual simulation in this field is carried out through three-dimensional animation production and display helmets, tactile simulation in this field is carried out through sensing mechanisms and gloves, sound simulation is carried out through sound production and sound effect cards, and dynamic simulation is carried out through mechanical control and transmission devices. Then, the human's response to these sensory stimuli is fed back to the main control mechanism, so as to generate a simulation of the new sensory model in real time. The civil engineering is closely related to people's lives, establishing a



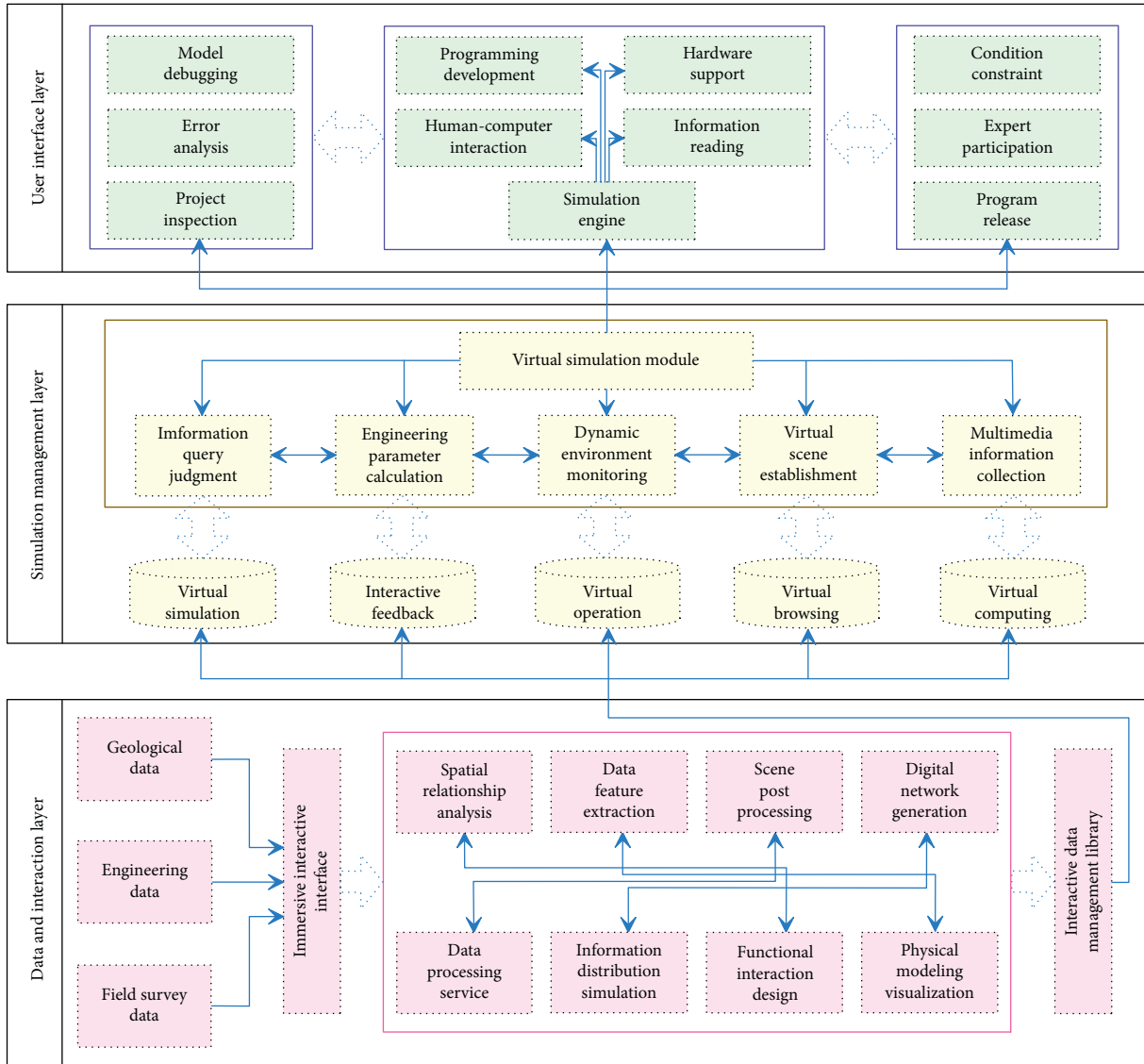


FIGURE 2: Framework and structure of virtual reality platform system.

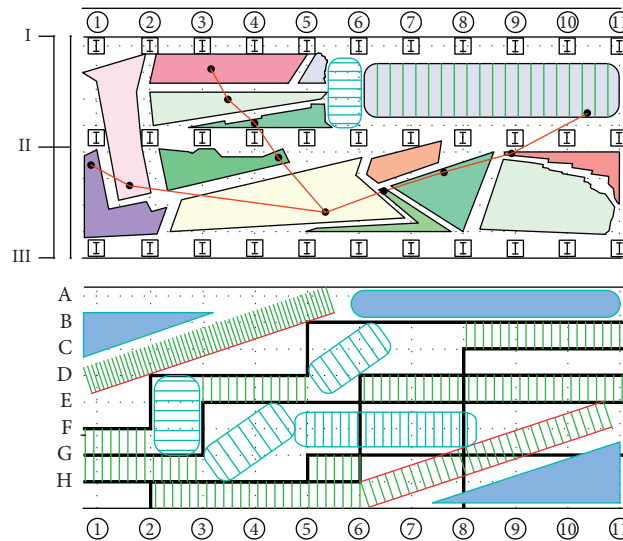


FIGURE 3: Planning display of civil engineering project based on virtual reality technology.

multidimensional information perception model and applying it to simulation systems is of great significance to the development of this industry [13].

**3.2. Synthetic Debugging and Roaming of Virtual Environment.** In the application process of the actual virtual simulation system, the display function can realize two-dimensional navigation, zoom-in, zoom-out, rotation, roaming, and corresponding animation functions; its query function can realize related query operations of attribute data, such as building name, area, weight Subordinate units, and completion time and can obtain corresponding target information from the database information. The simulation system actually records the city's topography, three-dimensional contours of buildings in key areas, and urban space, which can effectively help the relevant departments of urban planning, feel the construction status of the city more intuitively, and effectively evaluate it for comparison and optimization urban planning program. On the basis of effectively completing the modeling of virtual scenes, further synthesize and debug the virtual environment roaming system and this process requires the use of software to achieve the corresponding debugging effect (Figure 4). The debugging process is as follows: the hardware system and display device are firstly configured, and then project the screen, helmet, stereo glasses, etc. are installed, and set different operation modes, roaming paths, roaming viewing angles, running backgrounds, etc., to detect and debug collision effects and stereo effects. By clicking the mouse, users can effectively observe different angles in the virtual scene and adjust to watch the best effect.

Planning design is a very important link in the construction of civil engineering, and it is also an important foundation for civil engineering construction, which plays a decisive role in the successful completion of construction in the later stage of civil engineering and even affects the quality of the overall civil engineering. For civil engineering construction, the structure is very complex, and the construction environment has the characteristics of variability [14]. Designers usually spend a lot of energy and time to collect civil engineering construction data, and only after a comprehensive analysis of these data, we will be able to draw a complete civil engineering graphic design. The introduction of the VR technology can solve this problem well and designers can complete the mechanical performance model test through the computer, which avoids the influence of factors such as airflow and friction in the previous mechanical performance model test, which greatly improves the mechanical performance. The accuracy of the test and the test data can be comprehensively analyzed through the computer, and at the same time, it can help the designer to select the best civil engineering design plan. The VR system will also integrate the various links of civil engineering construction and clarify the relationship between the various construction links, so as to ensure that the civil engineering works are carried out step by step, and the construction is stopped because of the wrong connection of the construction links.

The application of the VR technology can also change the angle arbitrarily and observe from any point when performing visual calculation. At the same time, it can also use the human-computer interaction function provided by this technology to modify various calculation data in real time, so that make a detailed comparison of various schemes and results. Applying visual computing technology based on VR technology to the design, engineering control, and structural analysis of ultralarge and complex structures in civil engineering can effectively enhance the processing power of related visual computing software [15]. As far as the construction of civil engineering is concerned, the selection and optimization of engineering construction plans is an important task of engineering construction, and it is also an important guarantee to ensure the quality of engineering construction. In order to arrange the construction order, the application of this technology in the selection and optimization of construction plans can be used for the virtual construction and demonstration of the construction plans of various projects, and the construction plans can be compared through the virtual demonstration results, so as to effectively realize the selection and optimization of construction plans. When using the finite element method for structural analysis of the visual calculation of civil engineering, the application of the VR technology can mark the force of each point in the three-dimensional object with different shades of color, and use different colors to indicate Isobaric surface.

## **4. Planning Simulation of Civil Engineering Project Planning Based on VR Technology**

**4.1. Implementation of Virtual Construction Process and Structural Calculation.** The civil engineering planning needs to consider many factors, such as land use, economy, transportation, landscape, laws and regulations and other social and economic factors, meteorology, geology, terrain, hydrology, and other natural factors as well as water quality, noise, environmental pollution, greening and beautification, factors of living environment, etc. Planning information storage and query system, such as soil database system, regional information system, geographic information system, geographic information system, urban policy information system, and this type of system mostly uses the form of a database system. The shortcoming of the current database is the high degree of digitization and the low degree of visualization and these kinds of data are abstract and not easy to accept (Figure 5). For example, the geographic information system's representation of terrain and landforms is poorly readable if it is only represented by numbers. If it is represented in the form of a topographic map, it is relatively easy to accept, and the VR technology is used to input terrain; geomorphic data can be observed from different angles. Not only can the necessary data be obtained but also an intuitive experience can be obtained. Planning auxiliary performance integrated systems such as landscape performance system, and transportation planning system. At present, the representation method of landscape representation system is mainly two-dimensional pictures. If users

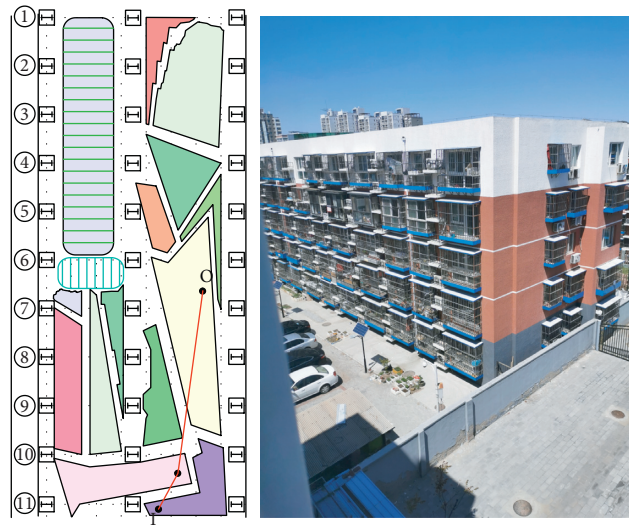


FIGURE 4: Planning display of civil engineering project from the perspective of O-I.



FIGURE 5: Planning display of civil engineering project from the perspective of O-II.

can feel a sense of immersion, landscape planning will be more scientific, reasonable, and comprehensive [16].

Construction plan design based on VR refers to the use of VR technology to establish a three-dimensional model of construction scenes, structural members, and mechanical equipment in a virtual environment to form a computer-based simulation system with certain functions and make the model in the system has dynamic performance and performs virtual construction on the model in the system. According to the result of virtual construction, it is verified whether it is correct. The construction plan is designed and modified in the visual environment of human-computer interaction to identify potential safety hazards and formulate safety precautions to get an optimized design plan. The identified hazards should be evaluated one by one by the construction company and the project manager. The major hazards identified can be combined with virtual environment research, detailed preventive measures, and feedback

to the hazard identification database. In the course of the project, the project manager department should adjust the identification of hazard sources, evaluate them, and formulate corresponding preventive measures at any time according to the safety conditions on-site, such as the implementation of safety plans, the accident conditions, and the generation of new hazard sources. The system then feeds back to the hazard source identification result database. After the hazard source identification results of the project are aggregated into the hazard source identification result database, the construction company can be used to continuously supplement and improve the standardized safety management system.

Improving the engineering measurement level can not only ensure the smooth progress of engineering construction and avoid construction changes during the construction process but also ensure that the quality of civil engineering construction meets the design requirements. In the

construction process of civil engineering, it usually involves three types of measurement work: angle measurement, distance measurement, and elevation measurement [17]. It is relatively tedious, and the measurement quality is easily affected by various factors and measurement errors. The application of VR technology in engineering survey work can build a virtual model by collecting environmental information, use computer technology to automatically measure based on the model, accurately analyze the measurement data, and issue a measurement report. As a result, multiple tasks such as measurement, data recording, data analysis, and drawing are combined in the same system to avoid investing too much manpower and material resources in the measurement work, while improving measurement accuracy and measurement efficiency. For example, in a high-rise residential building construction project, the surveyor needs to perform distance measurement, just input the engineering geographic environment data and residential building design data into the system, and build a three-dimensional model, and the system can accurately perform data measurement and analysis and intuitively present the measurement results on the three-dimensional model to facilitate the next work of the surveyor [18].

*4.2. Implementation of Virtual Project Management and Visualization.* As an emerging branch of civil engineering, engineering project management runs through the entire life cycle of construction projects. From project conception, planning, and design to project establishment, construction and other VR technologies provide more effective, more intuitive, and interactive engineering project management methods and approaches for engineering project management. The VR technology is a comprehensive high-tech information technology with a wide range of applications, which introduces its application in project feasibility studies and investment economic evaluation, real estate, bidding, construction management, property management, etc. and points out that this technology is used in civil construction with broad application prospects in project management (Figure 6). In actual engineering construction, the design of complex structure construction schemes and the calculation of construction structure are more difficult problems. The key to the former lies in the expression of the spatial relationship between the structural members and mechanical equipment on the construction site and the latter lies in the construction structure under construction. The deformation under state and load is greater than that after it is in place or after the structure is shaped. Especially modern large-scale engineering projects, long construction period, large quality, and even a series of enterprises produced by a project involve the coordination of various types of work, a lot of capital and material scheduling, construction machinery, and equipment management.

Visual processing or three-dimensional graphic display of the data obtained by scientific computing can be used to interactively change the parameters to observe the full picture of the calculation results and their changes and to achieve parameterized and visual calculations. When using

the finite element method for structural analysis, it can be used to give the force of each point in the three-dimensional object through the depth of the color; use different colors to represent different force surfaces. The users can also change the angle to observe by clicking on the interface and can also use the interactive performance of VR to modify various data in real time in order to compare various programs and results, which makes the engineer's thinking more visual and the concepts easier to understand [19]. Applying visual computing technology to the design, engineering control, and structural analysis of very large and complex structures will enhance the front-to-rear processing capabilities of computing software. More importantly, it can use graphics or images to re-analyze and dynamically control the structure in real time and obtain construction control data and at the same time can dynamically demonstrate and control the design and construction process in real time. The VR technology is constantly evolving and high-performance chips dedicated to computer graphics and multimedia information processing can increase the processing power by a hundredfold. Three-dimensional graphic algorithms and parametric modeling algorithms can make VR technology more mature.

Virtue reality has been tried to be applied in civil engineering construction and management. The use of its technology can simulate the construction site and construction activities, manage human and material resources in the construction organization, and identify hidden safety hazards. The traditional construction process management, organization, and safety plan are usually based on the provided design drawings. Since the drawings are two-dimensional ones provided by the designer, the understanding of these drawings requires a long translation process. Therefore, when planning, arranging, and managing construction, it is difficult to quickly form a global impression in the minds of engineering technicians and construction personnel. It is worth noting that the safety factors in construction often exist in the construction process, and it is impossible to find out from the drawings [20]. The use of VR technology can realize what users see is what you get and users can set the construction process parameters through interactive operation, analysis, and evaluation construction plan. In civil engineering, visual computing is an important trend in its future development and effectively integrating the VR technology can speed up the development of civil engineering visual computing technology to a large extent. The analysis of VR technology in urban planning and civil engineering should be analyzed, which is conducive to improving the overall transformation speed and construction quality of cities.

## 5. Simulation Experiment and Result Analysis

*5.1. Project Overview and Simulation Environment Design.* The V-Realm Builder visual editor provided by MATLAB greatly simplifies the modeling process, which is simple and easy to master, and the virtue reality modeling language is used to realize the dynamic visualization and interactive body of the virtual world. This modeling language is a scene



FIGURE 6: Planning display of civil engineering project from the perspective of O-III.

modeling language used to build real-world scene models or people's fictional three-dimensional worlds and is also platform-independent. It provides several degrees of freedom, can move in three directions, can rotate in three directions, and can also create hyperlinks in other spaces. The virtual world of the model created by the modeling language has all the static nodes of the three-dimensional world, including geometric relationships, backgrounds, textures, light, and viewpoints.

Two modeling methods are adopted for different objects: nonparametric modeling and parametric modeling. The nonparametric modeling method is only suitable for situations where there are a large number of identical shapes, and the size does not need to be changed. For relatively complex shapes and many dimensions, in order to install accurately, no deformation is allowed, and the size does not need to be changed, and then a nonparametric model is used. Parametric modeling is suitable for situations where the structure is more complex and there are more variant modeling constraints. The floor, staircase, tower crane, and other models are entered into the Envision module to form multiple institutions based on functional characteristics. In the Envision module, the mechanism is the smallest unit that can move independently. By positioning the organization, the entire organization model is assembled.

**5.2. Results Analysis.** The VR system can help planners organize and synthesize massive data in three-dimensional scenes. Based on the three-dimensional scene, the element information can be separately imported into the VR scene to form a database of all element space combinations, and planners can intuitively elaborate all element information. The VR technology is formed on the basis of true three-dimensional technology and can display all kinds of reality element information in VR scenes without additional text, symbols, and tables for annotation. After the application of VR technology to achieve results, the group evaluation and

modification phase is very convenient. When problems are found, all the modification work can be completed in real time in the VR environment. With the application of VR technology, during the review stage, the three-dimensional planning results can be displayed on the spot, so that the reviewers can feel the effects of the planning, experience the planning intention, and achieve the purpose of fully displaying the planning results (Figure 7). The results of VR can record the implementation of urban planning measures in real time. By modifying and recording the objects in the planned three-dimensional scene at any time, the current status documents of different periods are generated, and all the documents are stored as a database of the urban development process [21].

The reasonable choice of construction plan is the core of project organization design, which includes determining the construction flow direction and application procedures, selecting construction methods and construction machinery, and arranging construction sequence. For some projects with complex structure and large amount of engineering, it is difficult to choose the construction plan. However, through the VR system, the construction plan of each branch project can be virtually applied and demonstrated, which brings great convenience to the choices of construction plan [22]. As an emerging branch of civil engineering, engineering management runs through the entire life cycle of construction projects from project conception, planning, and design for project initiation and construction. In the entire project construction process, people have to use abstract concepts to express very rich content, such as plane drawings, section drawings, elevation drawings, and other plane graphics plus some prescribed symbols to represent three-dimensional buildings, using more abstract graphics and concise language to describe complex scenes to convey a lot of information. However, this kind of information processing and transmission method is affected by the occupation, knowledge structure, and understanding ability of the information receiver, so it is very inconvenient to

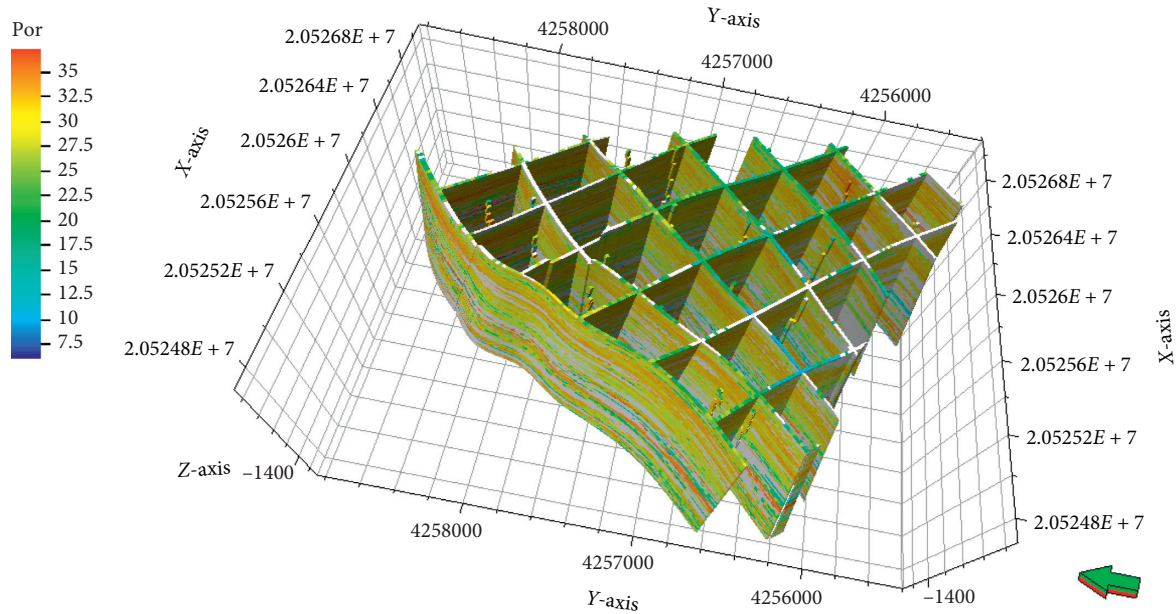


FIGURE 7: Panoramic display of civil engineering project based on virtual reality technology.

communicate. The development of the VR technology provides an extremely effective means to overcome this difficulty.

The computer simulation technology in the simulation technology realizes the constitutive relationship and mathematical modeling of civil engineering. For example, the constitutive relationship and failure criterion of concrete have many statements, and each failure criterion is only applicable under certain conditions within a certain range. In addition, the constitutive relation is based on the data obtained by the test, not derived from theory, and it is difficult to analyze its strength and tolerance in actual operation. The finite element method of computer simulation technology can be used to analyze and study concrete strength and other indicators. The simulation technology can also realize the full simulation of the design results, for example, building a simulation model of a bridge or other building through computer technology, then applying pressure or changing its structure through other means, and carrying out simulation force analysis under such changes [23]. Accurate results can be obtained, so that the mechanical characteristics of the building can be analyzed and studied. Through the application of computer technology, establish an information exchange platform within the civil engineering enterprise and improve the enterprise information sharing mechanism through the platform construction. Full communication and communication between the enterprise and employees can solve various problems in the construction of the project in a timely manner, promote the successful completion of civil engineering construction, and at the same time protect the two-way interests of the enterprise and employees.

## 6. Conclusions

This paper introduced the methods and principles of VR platform system and structure dynamic analysis, conducted virtual scene design and three-dimensional modeling of civil engineering project, analyzed the synthetic debugging and simulation roaming of the virtual environment, proposed the engineering model of virtual construction process and planning structure calculation, discussed the realization method of virtual project management and visualization, and finally performed the simulation and its result analysis of panoramic display and planning of civil engineering project based on VR technology. The application of the VR technology can simulate the construction site and construction activities, manage the human and material resources in the construction organization, and identify the construction safety hazards, which can also change the angle arbitrarily and observe from any point when performing visual calculation. From project conception, planning, and design to project establishment, construction and other VR technologies provide more effective, more intuitive and interactive engineering project management methods and approaches for engineering project management. Civil engineering project planning information storage and query system, such as soil database system, regional information system, geographic information system, geographic information system, and urban policy information system mostly use the form of the database system. The results show that the VR technology can simulate various environments and activities of civil engineering projects in advance, analyze the rationality of various schemes, and modify their different functional parameters, so it can achieve the simulation optimization of project planning, design, and construction. The VR technology can also use the three-dimensional

environment generated by electronic computers to make users feel the realistic existence of objects in virtual environment and also enable users to make accurate responses to movements and operations in the virtual world in a timely manner. The results of this study provides references for further researches on the panoramic display and planning simulation of civil engineering project based on VR technology.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this study.

## References

- [1] W. Zhu and G. Fan, "Application of computer virtual reality technology in virtual tour," *International Journal of Advanced Media and Communication*, vol. 6, no. 2–4, pp. 273–282, 2016.
- [2] R. S. Overbeck, D. Erickson, D. Evangelakos, M. Pharr, and P. Debevec, "A system for acquiring, processing, and rendering panoramic light field stills for virtual reality," *ACM Transactions on Graphics*, vol. 37, no. 6, pp. 1–15, 2018.
- [3] X. Lu and S. Davis, "Priming effects on safety decisions in a virtual construction simulator," *Engineering, Construction and Architectural Management*, vol. 25, no. 2, pp. 273–294, 2018.
- [4] A. Z. Sampaio, "Building maintenance supported on virtual reality environments: roofs and walls," *Journal of Civil Engineering and Architecture*, vol. 10, no. 1, pp. 1–12, 2016.
- [5] R. Napolitano, A. Blyth, and B. Glisic, "Virtual environments for visualizing structural health monitoring sensor networks, data, and metadata," *Sensors*, vol. 18, no. 1, p. 243, 2018.
- [6] Y.-C. Lin, Y.-P. Chen, H.-W. Yien, C.-Y. Huang, and Y.-C. Su, "Integrated BIM, game engine and VR technologies for healthcare design: a case study in cancer hospital," *Advanced Engineering Informatics*, vol. 36, pp. 130–145, 2018.
- [7] D. Hill, "Virtual reality headsets make their way into construction and design," *Civil Engineering Magazine Archive*, vol. 86, no. 1, p. 37, 2016.
- [8] L. A. Shuster, "At CES, drones and virtual reality devices target the construction industry," *Civil Engineering*, vol. 86, no. 2, pp. 42–43, 2016.
- [9] C. A. Cardno, "Virtual and augmented reality resolve remote collaboration issues," *Civil Engineering*, vol. 86, no. 9, pp. 40–43, 2016.
- [10] V. N. Glukhikh and N. V. Norina, "Experience of technical disciplines remote training at the st. Petersburg state university of architecture and civil engineering," *Education and Information Technologies*, vol. 21, no. 5, pp. 1401–1412, 2016.
- [11] Z. Zou, L. Arruda, and S. Ergon, "Characteristics of models that impact transformation of BIMs to virtual environments to support facility management operations," *Journal of Civil Engineering and Management*, vol. 24, no. 6, pp. 481–498, 2018.
- [12] D. Zhao, A. Mccoy, B. Kleiner, and Y. Feng, "Integrating safety culture into OSH risk mitigation: a pilot study on the electrical safety," *Journal of Civil Engineering and Management*, vol. 22, no. 6, pp. 800–807, 2016.
- [13] A. Bustillo, M. Alaguero, I. Miguel, J. M. Saiz, and L. S. Iglesias, "A flexible platform for the creation of 3d semi-immersive environments to teach cultural heritage," *Digital Applications in Archaeology and Cultural Heritage*, vol. 2, no. 4, pp. 248–259, 2015.
- [14] D. Fritsch and M. Klein, "3D preservation of buildings - reconstructing the past," *Multimedia Tools and Applications*, vol. 77, no. 7, pp. 9153–9170, 2018.
- [15] N. Tahani, A. A. Alesheikh, and M. Farnaghi, "Representation of 3D city models based on two interactive approaches in mobile devices," *Journal of Civil and Structural Engineering*, vol. 6, no. 2, pp. 145–160, 2016.
- [16] C. A. Sanchez, B. L. Ruddell, R. Schiesser, and V. Merwade, "Enhancing the t-shaped learning profile when teaching hydrology using data, modeling, and visualization activities," *Hydrology and Earth System Sciences*, vol. 20, no. 3, pp. 1289–1299, 2016.
- [17] J. Rubenstone, "Virtual reality finds a niche in skilled-trades education," *Engineering News Record*, vol. 279, no. 12, p. 31, 2017.
- [18] I. Freeman, J. Salmon, and J. Coburn, "A bi-directional interface for improved interaction with engineering models in virtual reality design reviews," *International Journal for Interactive Design and Manufacturing*, vol. 12, no. 6, pp. 1–12, 2017.
- [19] S. Choi, K. Jung, and S. D. Noh, "Virtual reality applications in manufacturing industries: past research, present findings, and future directions," *Concurrent Engineering*, vol. 23, no. 1, pp. 40–63, 2015.
- [20] F. Rumsey, "Virtual reality mixing, rendering, believability," *Journal of the Audio Engineering Society*, vol. 64, no. 12, pp. 1073–1077, 2016.
- [21] M. Sun, J. Pan, C. Le Bastard, Y. Wang, and J. Li, "Advanced signal processing methods for ground-penetrating radar: applications to civil engineering," *IEEE Signal Processing Magazine*, vol. 36, no. 4, pp. 74–84, 2019.
- [22] T. A. Adagunodo, A. A. Adeniji, A. V. Erinle, S. A. Akinwumi, and O. T. Kayode, "Geophysical investigation into the integrity of a reclaimed open dumpsite for civil engineering purpose," *Interciencia*, vol. 42, no. 11, pp. 324–339, 2017.
- [23] M. Moosazadeh, S. Kharkovsky, J. T. Case, and B. Samali, "Antipodal Vivaldi antenna with improved radiation characteristics for civil engineering applications," *Microwaves, Antennas and Propagation*, vol. 11, no. 6, pp. 796–803, 2017.

## Research Article

# Research of the Context Recommendation Algorithm Based on the Tripartite Graph Model in Complex Systems

Fei Long 

*School of Economics and Management, Changsha University, Changsha 410003, China*

Correspondence should be addressed to Fei Long; [z20120995@ccsu.edu.cn](mailto:z20120995@ccsu.edu.cn)

Received 17 June 2020; Revised 18 July 2020; Accepted 28 July 2020; Published 5 October 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Fei Long. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of information technology, the information overload has become a very serious problem in web information environment. The personalized recommendation came into being. Current recommending algorithms, however, are facing a series of challenges. To solve the problem of the complex context, a new context recommendation algorithm based on the tripartite graph model is proposed for the three-dimensional model in complex systems. Improving the accuracy of the recommendation by the material diffusion, through the heat conduction to improve the diversity of the recommended objects, and balancing the accuracy and diversity through the integration of resources thus realize the personalized recommendation. The experimental results show that the proposed context recommendation algorithm based on the tripartite graph model is superior to other traditional recommendation algorithms in recommendation performance.

## 1. Introduction

With the rapid development of the Internet and the increasing popularity of smart phones, the information that people can access was increasing in complex systems. It became difficult to obtain information and easy to get lost in a huge amount of information. The data included information of major websites and mobile APPs. If the data had given a certain treatment, the information could have been more user-friendly and more efficient. Meanwhile, users would have been provided with a friendly experience, and the information provider could have had a higher revenue. So, the recommendation system came into being [1–4]. The introduction of complex context information brings more ideas to improve the efficiency of the recommendation algorithm. The main complex context information includes user context information, such as users' ages, occupation, and region; physical context information, such as location, weather, and time. There is much complex context information that could be considered by the recommendation algorithm. The context information can bring more accurate recommendations to the users of the recommendation system for complex real-world applications [5–9].

By applying recommendation technology, users are no longer passive recipients of services, but become active participants, directly affecting the way and content of services. With the rapid development and popularization of mobile information services and the continuous advancement of data collection technology, the dimension of the information we can collect has been continuously improved [10]. Traditionally, only the “user-item” two-dimensional recommendation model considers the amount of information collected. Relatively few data can be obtained with relatively low data dimensions, but in the context of mobile information services, relying only on the “user-item” two-dimensional recommendation model cannot generate a user for a given situational context effective personalized recommendations [11–14]. Therefore, it is very important and urgent to provide personalized recommendations for mobile users in specific situations.

There are three commonly used context-aware recommendation algorithms, and they are contextual prefiltering, contextual postfiltering, and contextual modeling. In this paper, a context recommendation algorithm based on the tripartite graph is proposed for the three-dimensional model. Improving the accuracy of the recommendation by



the material diffusion, through the heat conduction to improve the diversity of the recommended objects, and balancing the accuracy and diversity through the integration of resources thus realize the personalized recommendation of users. This algorithm focuses on projects with high popularity and improves the accuracy of recommendation. The experimental results show that the proposed recommendation algorithm on the accuracy and recall evaluation index has obvious improvement in performance than the traditional recommendation algorithm.

The remainder of this paper is organized as follows: Section 2 introduces the related work. Section 3 is the description and theory of the context recommendation algorithm based on the tripartite graph model. Section 4 presents the experimental process and results. The summary of the related work of our proposed model and algorithm and the conclusion of the paper are given in Section 5.

## 2. Related Work

The most important recommendation algorithms are the collaborative filtering algorithm and the content-based recommendation algorithm. However, there are certain problems that are hard to be solved in the two basic algorithms. Therefore, accuracy of the single recommendation algorithm is difficult to improve. Many scholars have shifted their goals from traditional recommendation algorithms, focusing on the research of context-based recommendation systems.

The research direction and application fields include shopping, tourism, catering, and other aspects, but they focused on proposing improved algorithms which are based on traditional algorithms. The contrast of advantages and disadvantages of various algorithms based on context perception is relatively rare. The fields and data types applicable to different context-aware algorithms are not studied. Kang et al. proposed not only paying attention to the explicit context information, but also implicit context information. And, they used the packet sniffing technique [15]. Ohbyung Kwona and Jihoon Kim proposed a way to discover the user's portrait by depicting the user's outline tree [16]. Since 2012, related research has exploded. Dao et al. combined user preferences and contextual interaction information and proposed a location-based advertising recommendation algorithm [17]. Sánchez-Pi et al. proposed the construction of a knowledge-based context-aware system [18]. Kasaki et al. proposed a localized adaptive modification model using a vector space model and saved time to obtain user-related information [19]. Unger et al. made a breakthrough in the acquisition of contexts, improving the accuracy of recommendations by extracting implicit context information from users' data [20].

The contrast of advantages and disadvantages of various algorithms based on context perception is relatively rare. Before 2016, there were not many research studies on context-aware recommendation algorithms. Most of them focused on the research of collaborative filtering algorithms. Such research is rare. There are three commonly used context-aware recommendation approaches, and they are

contextual prefiltering, contextual postfiltering, and contextual modeling [21]. Among them, Lee et al. proposed a novel TV system that combines gesture control, tag ranking, and context awareness, providing personalized recommendations [22]. Cai et al. used collaborative filtering to generate contextual concept ontology [23]. Ji et al. proposed an improved matrix approximate based on the clustering model [24]. Wang et al. proposed a hierarchical recommendation model based on the context tree structure, calculated situation transfer in the first layer, and recommended items in the second layer [25].

In research on the context recommendation algorithm, a lot of scholars have contributed to it. Among the above existing literature studies, some only use the two-dimensional model to calculate the recommendation list. Most of the papers on the research of 3D models focus on the label-based 3D model. The research aspect is relatively narrow and limited [26–30]. In this paper, the context recommendation algorithm based on the tripartite graph model is introduced. Research is more innovative and has improved accuracy in recommendations.

## 3. User-Item-Context Tripartite Graph Model

*3.1. Tripartite Graph Model.* The context incorporation task is the process of managing and interpreting the modeled context to be directly integrated in the recommendation system. This step consists of identifying the user and his/her preferences extracted from his/her implicit and explicit interactions with the system. These preferences are sensitive to the user's context and the change from a context to another. Then, the so-called contextual profile, the user's preferences in a given context, can be defined by the tuple (user, preferences, and context). These tuples can be saved in a database and have to be regularly updated to keep track of the evolution of the user's interests.

The application of context-aware algorithms is based on context acquisition. Current technology of shallow context acquisition can already be achieved. The user's latitude and longitude position, weather conditions, time information, etc., can be obtained through mobile devices. Age, gender, and other information, but deep-level information processing is currently relatively small, such as obtaining the user's location type, time type, and deep-level context information, have a greater effect on algorithm recommendation. The current research on this type of information is mainly obtained through user surveys and is provided voluntarily by users.

The context plays an important role in the user's selection of the item. The interest of the user will also migrate due to the context. If the context is not introduced, the interest changes in different contexts are not observed. In different contexts, users may feel different for the same item. For example, people will watch action movies with friends, but they may choose other types of movies when they are with family. The traditional algorithm ignores these interest migrations, and the recommendation is not personalized enough. The three-dimensional model formed in the context

introduced in this paper is more effective in improving the limitations of the recommendation algorithm.

Among the basic graph theory algorithm, the bipartite graph is commonly used in the related network model. In the bipartite graph, there are two sets of sets  $U$  and  $I$ :

$$U = \{u_1, u_2, u_3, \dots, u_i\}, \quad (1)$$

$$I = \{i_1, i_2, i_3, \dots, i_j\}. \quad (2)$$

The recommendation of the item is realized by the bipartite graph model. We should ensure that the intersection between the two sets is an empty set and is not affected by each other. Each element in the two sets is treated as a different point in the graph, representing a two-dimensional model of the user item. The elements between the same set are not connected by edges, and the edges between different sets are connected indicating that the user likes the project, forming a binary group  $\langle u_i, i_j \rangle$ . That constitutes an undirected bipartite graph model. The bipartite graph model is shown in Figure 1.

The tripartite graph model is an upgraded version of the bipartite graph model and can be viewed as a merged version of two bipartite graphs. It includes three sets  $U$ ,  $I$ , and  $C$ , where the set  $C = \{c_1, c_2, c_3, \dots, c_k\}$ . Each set ensures that the two-two intersections are empty sets, so that the points of the three sets in the figure can be obtained. Similar to the bipartite graph model, the collections are guaranteed to be bottomless. There are two sets of two-tuple relations in the figure. There are two sets of edges connected, which are  $\langle u_i, i_j \rangle$  and  $\langle i_j, c_k \rangle$ . The former indicates that the user likes the project, and the latter indicates that the project is in a certain context. We can realize the recommendation through the three-tuple relationship. The tripartite graph model is shown in Figure 2.

Through Figures 1 and 2, the concept of the model is intuitively felt. The item acts as a transfer station, communicates with the user's information, communicates with the context information, and finally the recommended list of items is a recommendation algorithm. It is more comprehensive compared with the traditional algorithm.

### 3.2. Tripartite Graph Model Based on Mass Diffusion.

When the bipartite graph is used to obtain the recommended result, the item is associated with the user to obtain the order of the item's preferences. The commonly used algorithms are the mass diffusion algorithm and the heat conduction algorithm. It can be understood from the existing literature that the mass diffusion algorithm is energy conservation while the heat conduction algorithm is not. The two algorithms play different roles in the field of recommendation algorithms. The mass diffusion algorithm is usually likened to a convex lens, which can focus on items with high heat. The heat conduction algorithm is like a concave lens, which can focus on the item which is more unpopular, thus increasing the diversity of recommendations.

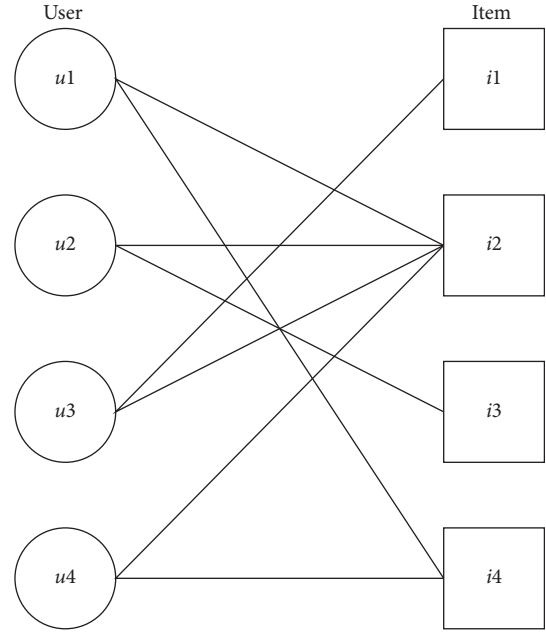


FIGURE 1: Bipartite graph model.

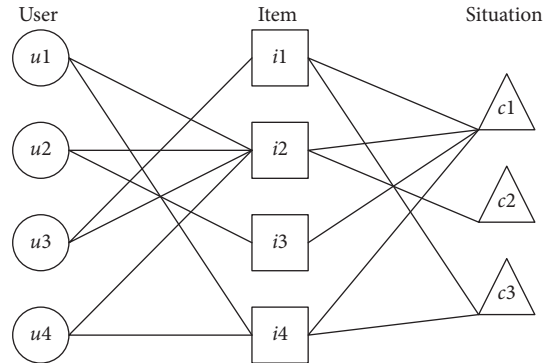


FIGURE 2: Tripartite graph model.

Among the results obtained, the nearest neighbor of each user is found, assuming that the choice of these nearest neighbors is related to the user's own choice. After finding the score of the target user  $u$  for the item  $i$ , the score of the most neighboring user for the item  $i$  is also found. Then, the mean of the nearest neighbors' score is calculated. If the mean value is bigger than the predicted score and the absolute value differs by more than a threshold value, the result is the predicted score adding the quotient of mean and predicted values. If the mean is equal to the prediction result or the absolute value difference is less than a threshold, no operation is performed. If the mean is smaller than the predicted score, the result is the predicted score subtracting the quotient of mean and predicted values.

The energy of each item is shown in the following equation:

$$c_{i,j} = \begin{cases} 1, & \text{user likes it,} \\ 0, & \text{user does not like it.} \end{cases} \quad (3)$$

According to the historical behavior of the user, the user's preference for the item is obtained. One means that the user likes the item, and zero means dislike. The total number of items preferred by the user is the total energy in the algorithm.

The first step of diffusion is from the item to the user. The user obtains energy from the item, and the item spreads the energy to the user by equally distributing the energy of each item to each purchased user. The energy obtained by the user is shown in the following equation:

$$b_\alpha = \sum_{l=1}^n a_{\alpha l} \left( \frac{c_l}{k(O_l)} \right), \quad (4)$$

in which  $k(O_l) = \sum_{i=1}^m a_{il}$  represents the degree of the item and  $c_l$  represents the initial energy possessed by the item. The energy value owned by each user is obtained by one-step diffusion, and then the second step is performed.

The second step of diffusion is from the user to the item, which distributes the user's energy equally to the related items, and each item sums the collected energy to obtain the total energy, as shown in the following equation:

$$c'_j = \sum_{\alpha=1}^m a_{\alpha j} \left( \frac{b_\alpha}{k(U_\alpha)} \right), \quad (5)$$

in which  $k(U_\alpha) = \sum_{j=1}^n a_{\alpha j}$  represents the degree of the user and  $b_\alpha$  is the energy obtained by the user as described above. One of the recommended coefficients of the model is obtained by averaging the energy obtained in the first step by the second material diffusion.

In the bipartite graph, the elements in the figure have been replaced, but the idea of the bipartite method has not changed. Two-step diffusion is used to generate the recommended probability. The first step is to spread from the item to the context. The initial energy is related to whether the user likes the item. One means that the user likes the item, and zero means dislike. According to the relationship between the context and the item, the energy of the item is evenly distributed. For each occurrence of the context, the energy obtained by each context is the sum of the energy allocated to all related items.

$$d_\beta = \sum_{h=1}^p a_{\beta h} \left( \frac{c_h}{k(S_\beta)} \right), \quad (6)$$

in which  $k(S_\beta) = \sum_{k=1}^q a_{\beta k}$  represents the degree of the context,  $c_h$  represents the initial energy possessed by the item, and  $c_h$  and  $c_j$  correspond to each other. The equation represents the energy of each context. The energy value of the context is obtained by the diffusion of the first step, and then the second step is still required to spread.

The second step is from the context of the item. Like the first step, the energy obtained by the context is divided into all the items, and the energy of each item is accumulated and obtained:

$$c'_t = \sum_{\beta=1}^q a_{\beta t} \left( \frac{d_\beta}{k(F_t)} \right), \quad (7)$$

in which  $k(F_t) = \sum_{i=1}^p a_{\beta t}$  represents the degree of the item and  $d_\beta$  represents the energy value obtained by the context. In the second diffusion, the average value of the energy obtained from the previous diffusion is obtained.

It can be seen from Figure 3 that by processing of the tripartite graph model based on mass diffusion, the recommendation value of the items can be calculated and the recommendation sequence can be obtained. The recommendation sequence is  $\{i_3, i_2, i_5, i_1, i_4\}$ .

### 3.3. Tripartite Graph Model Based on Heat Reduction.

This section discusses the heat conduction method. As mentioned earlier, the heat conduction method is a divergent method that recommends unpopular items for users to increase the diversity of recommendations. Like the tripartite graph model based on mass diffusion, this model is divided into two bipartite graphs, and two heat conduction methods are used, respectively. Finally, the values of the two graphs are fused to obtain the coefficients recommended by the model.

The energy of each item is in the same way as mentioned above:

$$c_{i,j} = \begin{cases} 1, & \text{user likes it,} \\ 0, & \text{user does not like it.} \end{cases} \quad (8)$$

According to the user's historical behavior, the user's preference for the item is obtained. One means that the user likes the item, and zero means dislike. The total energy in the heat conduction method is not conserved.

Similarly, the first step of conduction is from the item to the user. The user obtains the temperature of the item, and the item transmits the temperature of the user. The method of transferring the temperature is to divide the sum of the temperature of the user's favorite item by the degree of the user, and the temperature obtained by the user is as the following equation:

$$b_\alpha = \sum_{l=1}^n \frac{a_{\alpha l} c_l}{k(U_\alpha)}, \quad (9)$$

in which  $k(U_\alpha) = \sum_{j=1}^n a_{\alpha j}$  represents the degree of the user and  $c_l$  represents the initial temperature possessed by the item. The temperature possessed by each user is obtained by the first step of conduction, followed by the second step of conduction.

The second step of conduction is from the user to the item. The temperature of each item is the sum of the temperatures of all users who like it divided by the degree of the item:

$$t'_j = \sum_{\alpha=1}^m \frac{a_{\alpha j} b_\alpha}{k(O_j)}, \quad (10)$$

in which  $k(O_j) = \sum_{i=1}^m a_{ij}$  represents the degree of the user and  $b_\alpha$  is the temperature obtained by the user as described above. Through the second step of heat conduction, the average temperature value is obtained; that is, the obtained

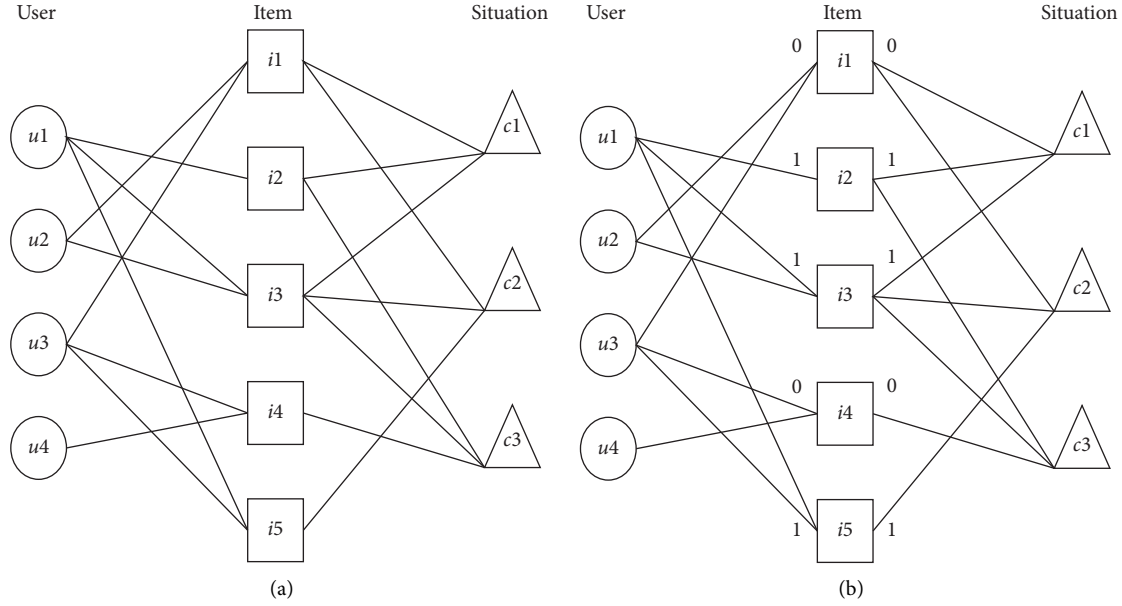


FIGURE 3: Processing of the tripartite graph model based on mass diffusion.

particle temperature value is one of the recommended coefficients of the heat conduction method.

The tripartite graph division method based on the heat conduction method is consistent with the material diffusion. For the item-context bipartite graph, the temperature of the calculated context is the sum of the associated item temperatures divided by the context:

$$d_{\beta} = \sum_{h=1}^p \frac{a_{\beta h} c_h}{k(F_t)}, \quad (11)$$

in which  $k(F_t) = \sum_{t=1}^p a_{\beta t}$  represents the degree of the context,  $c_h$  represents the initial temperature possessed by the item, and  $c_h$  and  $c_j$  correspond to each other. The equation represents the temperature of each context. The temperature of the context is obtained by the first step of conduction, and the second step is to conduct the temperature on the article.

The second step is to get the temperature of the item from the context of the item and divide the temperature of the context associated with the item with the degree of the item:

$$t'_k = \sum_{\beta=1}^q \frac{a_{\beta k} d_{\beta}}{k(S_{\beta})}, \quad (12)$$

in which  $k(S_{\beta}) = \sum_{k=1}^q a_{\beta k}$  represents the degree of the item and  $d_{\beta}$  represents the temperature obtained by the context. In the second conduction, the energy value obtained from the previous diffusion is averaged to obtain another recommendation coefficient.

Processing of the tripartite graph based on heat conduction is shown in Figure 4. Processing of the tripartite graph based on heat conduction is that nodes absorb resources from their adjacent nodes according to the number of adjacent nodes and the ratio of edge weight. After the

tripartite graph is divided into two bipartite graphs, the two recommended coefficients are obtained. In order to obtain the final recommendation coefficient, the average of the recommended coefficients obtained by the two is taken as the final recommendation coefficient:

$$E' = \frac{t'_j t'_k}{2}. \quad (13)$$

**3.4. Mixed Model.** The mass diffusion method and the heat conduction method are different in the recommendation system to improve the recommendation effect. The mass diffusion method focuses on improving the accuracy of the recommendation and the items with high popularity. The heat conduction method focuses on improving the diversity of recommendations. Therefore, the combination of the two methods is also a factor that is often considered, with the goal of ensuring accuracy and increasing the diversity of recommendations.

When mixing the two methods, it is usually a linear blending. We set the parameter  $\lambda$ , introducing parameters on the original three-part graph model, and construct a hybrid model:

$$e = \lambda E + (1 - \lambda) E'. \quad (14)$$

It is a common method in the hybrid model to change the value of the parameter  $\lambda$  to achieve the optimal effect of the model. In this paper, the influence of different parameter values on the model will also be explored.

## 4. Experiment Analysis

The hardware environment of this experiment is Intel Core i5 processor, 8G memory. The software environment is 64-bit Windows operating system, VMware Workstation Pro

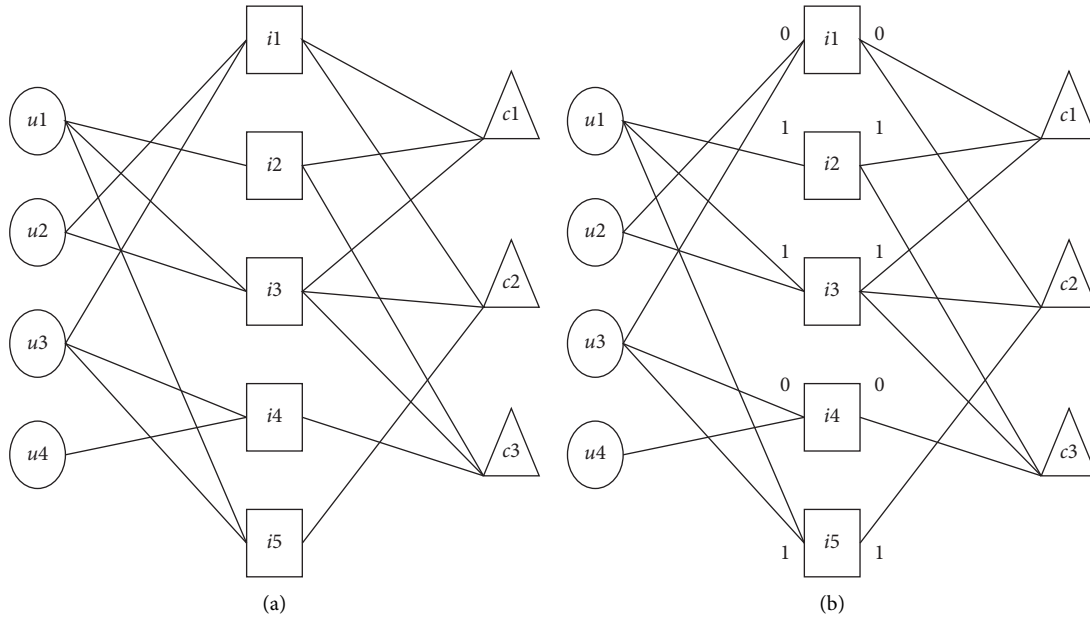


FIGURE 4: Processing of the tripartite graph based on heat conduction.

Virtual Machine, CentOS7 operating system, Open JDK1.8, Python3.7, and Anaconda for Python3.7.

The data of this experiment are mainly from the questionnaire survey of college students' movie-watching context. A total of 107 users' data are collected. In order to ensure the accuracy of the data, they are not excessively sparse. Some invalid questionnaires are eliminated. The data of 96 users were retained, including 2145 data.

In this experiment, the original data set is about the user's rating of the item, and the score range is 1–10 points. User's rating reflects the user's satisfaction with the item. The recommendation algorithm should refer to this satisfaction.

As shown in Figure 5, user A and user B scored item 1, but the score is completely different. So, the similarity between users is very low. Therefore, when the tripartite graph model is established, only the items that the user likes are selected. The connection line between the user and the item indicates that the user prefers an item above a certain rating range.

The evaluation parameters used in this experiment are a standard for evaluating the quality of the recommended results. The precision of the recommended results and the recall rate are used. Precision measures the proportion of recommended items that result relevant to the users, that is, those recommended items that the user actually consumes. Recall measures the proportion of consumed items that were correctly recommended; that is, the fraction of items relevant to the user that was suggested by the system. Recall and precision are usually considered together as two facets of the quality of the recommendation. The recall and precision are given as follows:

$$\text{precision} = \frac{\text{hits}}{\text{items\_relevant}}, \quad (15)$$

$$\text{recall} = \frac{\text{hits}}{\text{items\_recommended}}, \quad (16)$$

where *hits* are the number of correct recommendations, *items\_relevant* is the number of all possible items that can be recommended for a given user, and *items\_recommended* is the number of items recommended to a user. The precision and the recall rate are between 0 and 1. The closer it is to 1, the better the effect of the recommendation algorithm is. Closer to 0 indicates that the recommended effect of the recommendation algorithm is very poor.

This article also uses novelty to judge the novelty of an item:

$$\text{novelty} = \frac{\sum_{i=1}^m p(i)}{l}, \quad (17)$$

in which  $p(i)$  represents the degree of the item in the recommendation list and  $l$  represents the length of the recommendation list. In this paper, the average popularity is used to represent the novelty, and the smaller the value, the more novel the item.

The original data of this experiment are about the user's score on the movie in some contexts. It is converted into the connection between the user and the item, the item and the context, and the connection between the user and the item. We establish the connection between the item and the context, and the form is shown in Table 1.

If the users give the item with rate greater than or equal to 8, we consider the users to be very fond of the items. Based on this rule to establish the connection between the users and the items, as well as the connection between items and context, the total number of filtered data is 1451. Some relevant predictions are made for these data. For this experiment, two control experiments were set up to determine the precision and recall rate under the collaborative filtering algorithm and the user-item bipartite graph algorithm and compare the results of the algorithm.

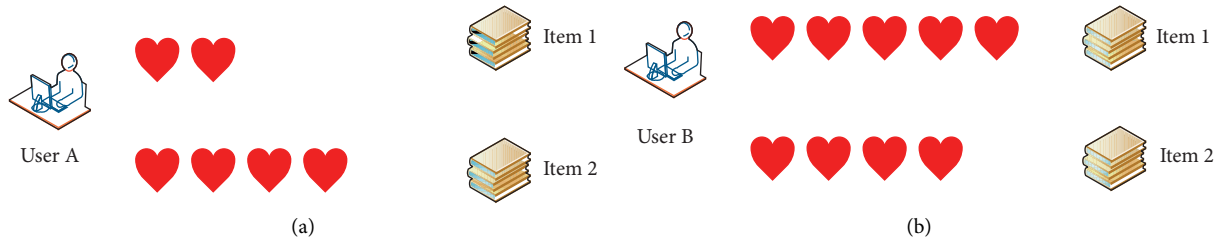
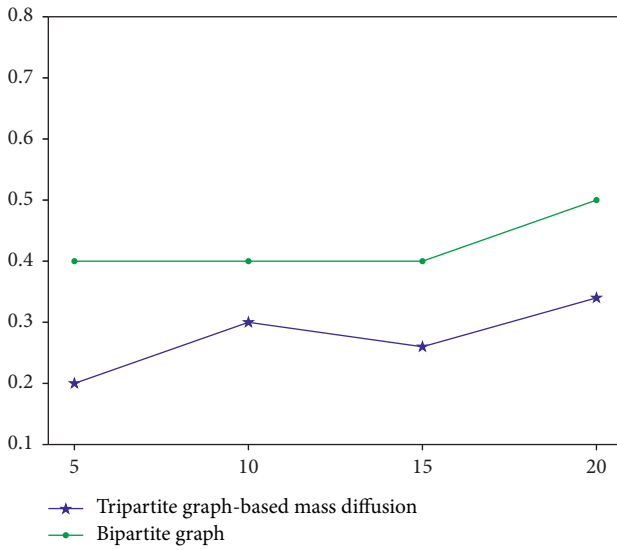


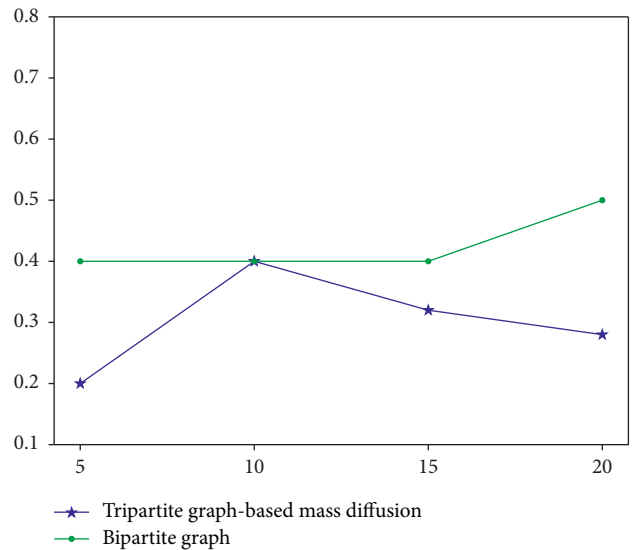
FIGURE 5: The score of items.

TABLE 1: The items that users prefer (part).

|    | User_id | Movie_id | S_id | Score |
|----|---------|----------|------|-------|
| 0  | 1101    | 2102     | 20.0 | 8.0   |
| 3  | 1101    | 2110     | 13.0 | 8.0   |
| 4  | 1101    | 2113     | 10.0 | 10.0  |
| 5  | 1101    | 2115     | 10.0 | 10.0  |
| 6  | 1101    | 2116     | 5.0  | 8.0   |
| 8  | 1101    | 2120     | 13.0 | 9.0   |
| 9  | 1101    | 2121     | 13.0 | 8.0   |
| 11 | 1101    | 2125     | 7.0  | 9.0   |
| 17 | 1101    | 2132     | 13.0 | 9.0   |
| 19 | 1101    | 2135     | 21.0 | 9.0   |
| 20 | 1101    | 2136     | 13.0 | 8.0   |
| 21 | 1101    | 2137     | 19.0 | 8.0   |



(a)



(b)

FIGURE 6: Continued.

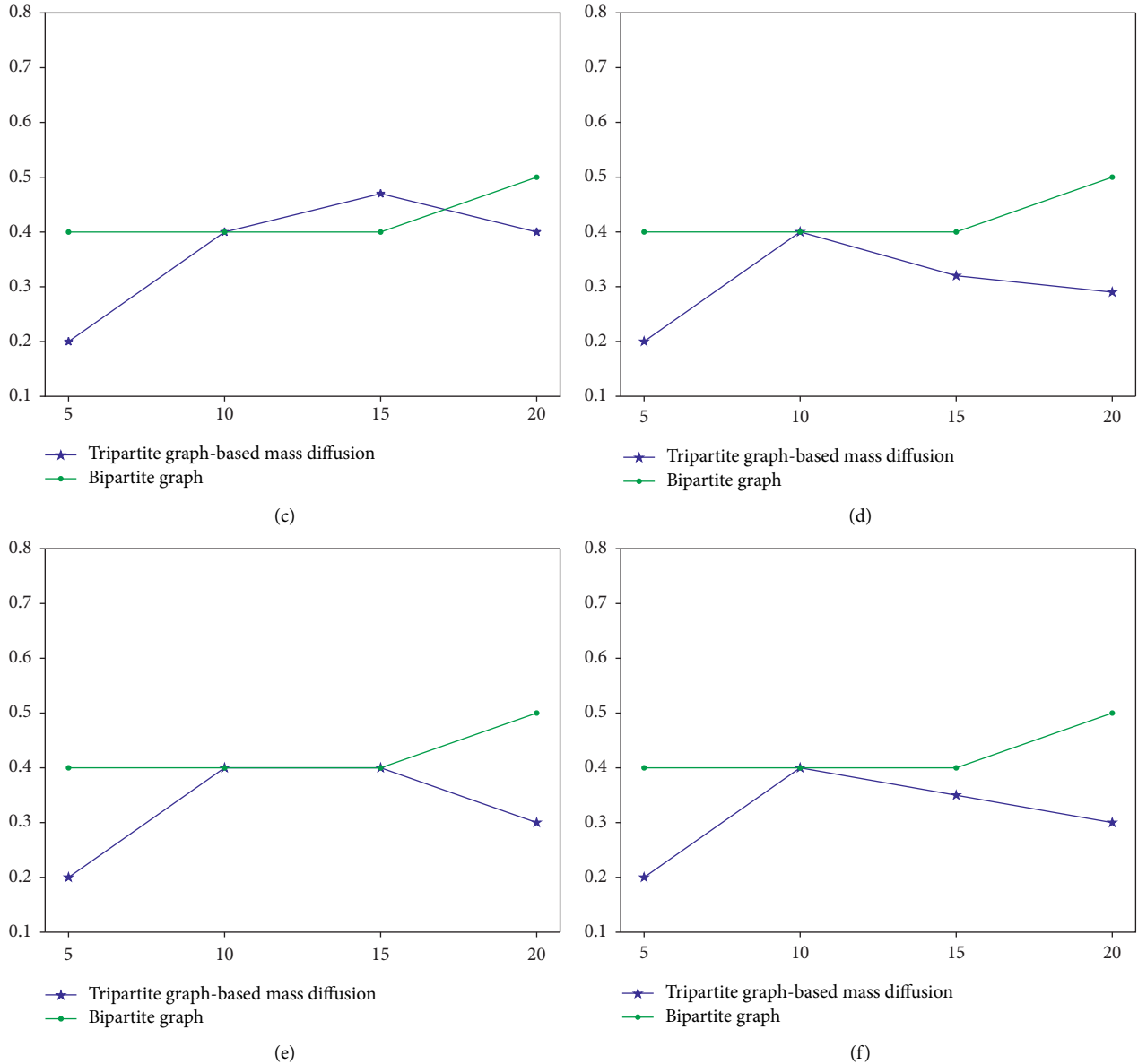


FIGURE 6: Precision of the hybrid model algorithm.

In the experiment, the user 1101 is taken as an example. The movie is recommended to be selected under the contexts 1, 5, 9, 13, 17, and 21, and 5, 10, 15, and 20 different movies are recommended, respectively. The comparison result of the accuracy rate and the recall rate is shown in Figures 6 and 7.

We calculate the novelty of the recommendation. The calculation base of the novelty is 20 when the recommended item is as shown in Table 2.

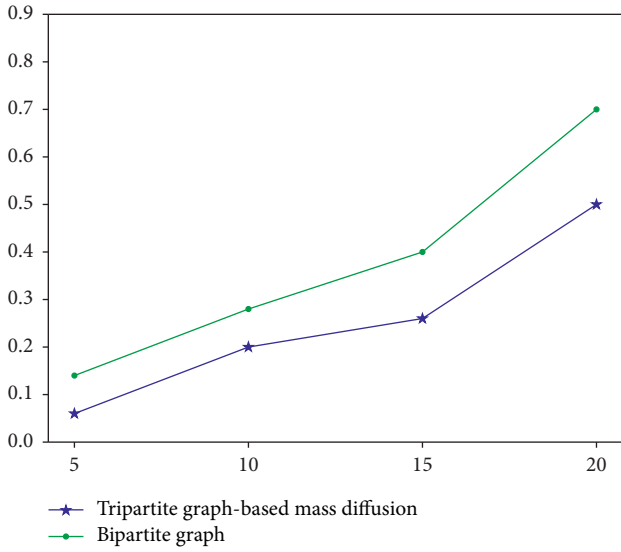
It can be seen from the comparison of the experimental results that in the experimental data set, the performance of the algorithm is not excellent enough. The precision rate is less reduced, and the recall rate is also somewhat reduced. When the recommended number of items is 10, the precision and recall rate of the algorithm are optimal.

Meanwhile, the novelty has also improved. Compared with the bipartite map recommendation algorithm, the items recommended by the three maps are more novel.

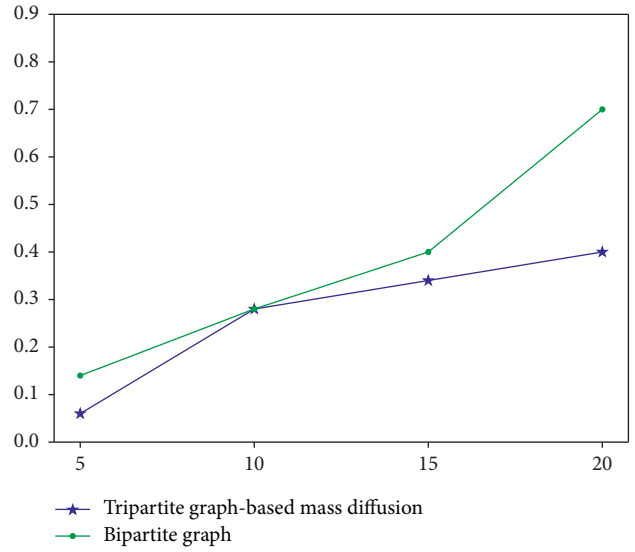
After the experiment, the precision and recall rate of the tripartite graph model based on heat conduction are almost zero, and the novelty is very high. Due to the divergence of the algorithm, the items focused on the unpopular, and the recommended results are not in line with actual needs, nor suitable for this data set.

In the hybrid algorithm, several of the mixed parameters are selected, which are 0.1, 0.3, 0.5, 0.7, and 0.9. We calculate the accuracy, recall, and novelty for each result as shown in Table 3.

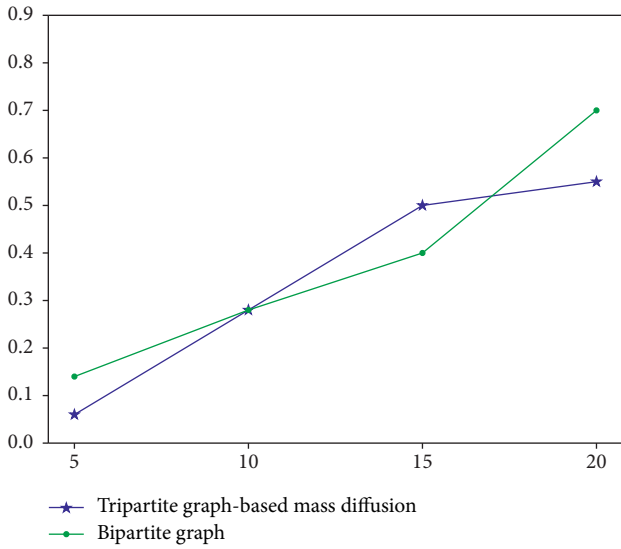
The performance of the hybrid model algorithm has basically stabilized when the value of  $\lambda$  is greater than 0.5.



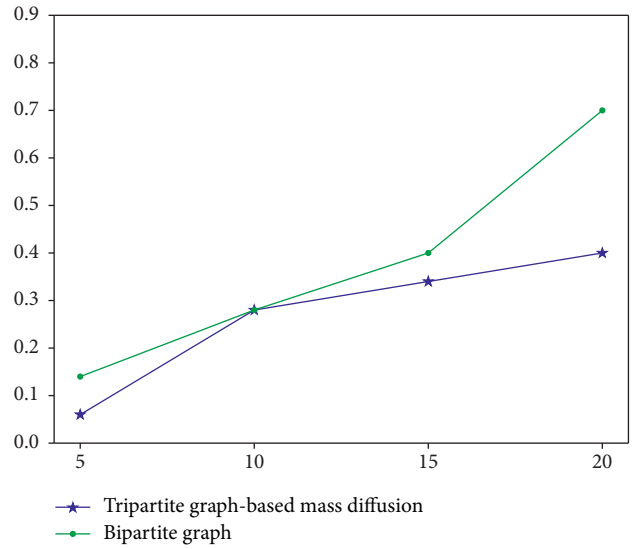
(a)



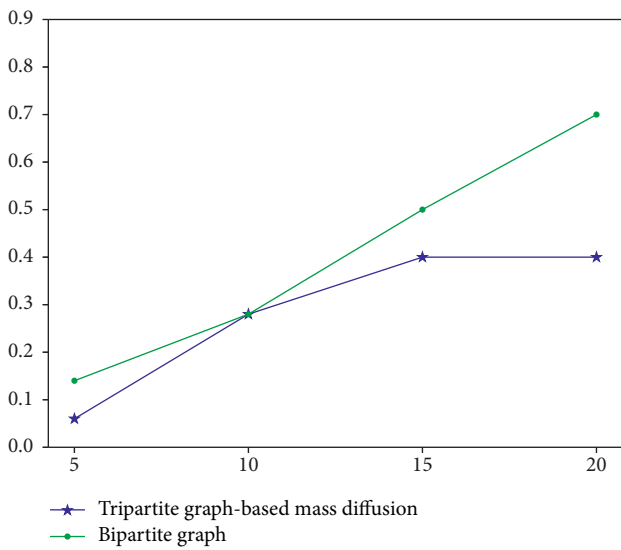
(b)



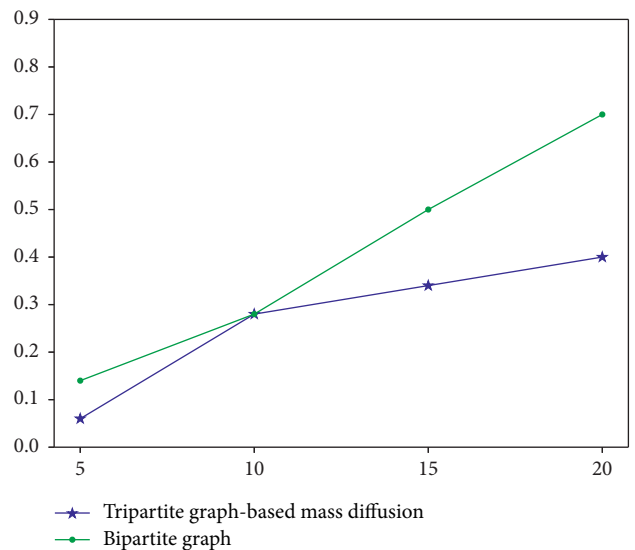
(c)



(d)



(e)



(f)

FIGURE 7: Recall of the hybrid model algorithm.



TABLE 2: Comparison of novelty.

|                  | 1     | 5     | 9    | 13    | 17    | 21    |
|------------------|-------|-------|------|-------|-------|-------|
| Bipartite graph  |       |       |      | 42.15 |       |       |
| Tripartite graph | 41.65 | 39.05 | 40.5 | 39.05 | 41.35 | 39.85 |

TABLE 3: Mixed model.

|           | 0.1    | 0.3    | 0.5   | 0.7   | 0.9   |
|-----------|--------|--------|-------|-------|-------|
| Precision | 0.1    | 0.2    | 0.3   | 0.3   | 0.3   |
| Recall    | 0.0714 | 0.1428 | 0.214 | 0.214 | 0.214 |
| Novelty   | 15.4   | 28.05  | 34.45 | 39.05 | 39.05 |

The change of parameters has little change on the algorithm result. The novelty shrinks as the parameter increases, and the accuracy increases as the parameter increases.

## 5. Conclusion

In this paper, to solve the problem of complex contexts, a new context recommendation algorithm based on the tripartite graph model is proposed for the three-dimensional model in complex systems. This algorithm focuses on projects with high popularity and improves the accuracy of recommendation. The range of recommended results is very narrow. Part of the reason is that the number of items in the data set is not large, and the recommendation results are not obvious. There is an urgent need to improve. And, the most popular items should be given some punishment mechanisms to enable more user interest. However, the novelty of the algorithm is slightly improved, and the performance of the algorithm tends to be stable after the model parameters reach a certain value [5–14, 26–30].

## Data Availability

The data used to support the findings of this study are available upon request from the corresponding author.

## Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was supported by the National Social Science Foundation of China (No. 18CTQ030).

## References

- [1] P. Wang, H. Luo, M. S. Obaidat, and T. Y. Wu, "The Internet of things service recommendation based on tripartite graph with mass diffusion," in *Proceedings of the IEEE International Conference on Communications Workshops*, May 2018.
- [2] B. Shams and S. Haratizadeh, "Reliable graph-based collaborative ranking," *Information Sciences*, vol. 432, pp. 116–132, 2018.
- [3] Y. Zhang and X. Chen, "Explainable recommendation: a survey and new perspectives," 2018, <https://arxiv.org/abs/1804.11192>.
- [4] W. Zhou and W. Han, "Personalized recommendation via user preference matching," *Information Processing & Management*, vol. 56, no. 3, pp. 955–968, 2019.
- [5] D. Das, D. Pochugari, N. Verma et al., *Suggesting Follow-Up Queries Based on a Follow-Up Recommendation Machine Learning Model*, Splunk Technology, San Francisco, CA, USA, 2019.
- [6] Z. Huang, X. Xu, J. Ni, H. Zhu, and C. Wang, "Multimodal representation learning for recommendation in Internet of Things," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10675–10685, 2019.
- [7] C. Musto, P. Basile, and G. Semeraro, "Hybrid semantics-aware recommendations exploiting knowledge graph embeddings," in *Proceedings of the International Conference of the Italian Association for Artificial Intelligence*, pp. 87–100, Rende, Italy, November 2019.
- [8] W. Zhou and W. Han, "Personalized recommendation via user preference matching," *Information Processing & Management*, vol. 56, no. 3, pp. 955–968, 2019.
- [9] Z. Shuang, B. Sheng, and S. Gengxin, "Mass diffusion recommendation algorithm based on multi-subnet composited complex network model," in *Proceedings of the International Conference on Artificial Intelligence and Security*, pp. 263–274, New York, NY, USA, July 2019.
- [10] C. Bernardis, M. F. Dacrema, and P. Cremonesi, "A novel graph-based model for hybrid recommendations in cold-start scenarios," 2018, <https://arxiv.org/abs/1808.10664>.
- [11] M. K. Najafabadi, A. Mohamed, and C. W. Onn, "An impact of time and item influencer in collaborative filtering recommendations using graph-based model," *Information Processing & Management*, vol. 56, no. 3, pp. 526–540, 2019.
- [12] D. Jiang, Z. Zheng, G. Li et al., "Gesture recognition based on binocular vision," *Cluster Computing*, vol. 22, pp. 13261–13271, 2019.
- [13] S. C. Lee, S. W. Kim, S. Park, and D. K. Chae, "A tripartite-graph based recommendation framework for price-comparison services," *Computer Ence and Information Systems*, vol. 16, no. 2, pp. 333–357, 2019.
- [14] R. Dridi, L. Tamine, and Y. Slimani, "Context-aware multi-criteria recommendation based on spectral graph partitioning," in *Database and Expert Systems Applications. DEXA 2019. Lecture Notes in Computer Science*, vol. 11707, Cham, Switzerland, Springer, 2019.
- [15] H. Kang, E. Suh, and K Yoo, "Packet-based context aware system to determine information system user's context,"

- Expert Systems with Applications*, vol. 35, no. 1-2, pp. 286–300, 2018.
- [16] O. Kwon and J. Kim, “Concept lattices for visualizing and generating user profiles for context-aware service recommendations,” *Expert Systems with Applications*, vol. 36, no. 2, pp. 1893–1902, 2009.
- [17] T. H. Dao, S. R. Jeong, and H. Ahn, “A novel recommendation model of location-based advertising: context-aware collaborative filtering using GA approach,” *Expert Systems with Applications*, vol. 39, no. 3, pp. 3731–3739, 2011.
- [18] N. Sanchez-Pi, J. Carbo, and J. M. Molina, “A knowledge-based system approach for a context-aware system,” *Knowledge Based Systems*, vol. 27, pp. 1–17, 2012.
- [19] N. Kasaki, S. Kurabayashi, and Y. Kiyoki, “A geo-location context-aware mobile learning system with adaptive correlation computing methods,” *Procedia Computer Science*, vol. 10, pp. 593–600, 2012.
- [20] M. Unger, A. Bar, B. Shapira, and L. Rokach, “Towards latent context-aware recommendation systems,” *Knowledge Based Systems*, vol. 104, pp. 165–178, 2016.
- [21] K. M. Kim, D. Kwak, H. Kwak et al., “Tripartite heterogeneous graph propagation for large-scale social recommendation,” 2019, <https://arxiv.org/abs/1908.02569>.
- [22] W.-P. Lee, K. Che, and J.-Y. Huang, “A smart TV system with body-gesture control, tag-based rating and context-aware recommendation,” *Knowledge Based Systems*, vol. 56, pp. 167–178, 2013.
- [23] W.-h. Chen, Y. Cai, and H.-f. Leung, “Generating ontologies with basic level concepts from folksonomies,” *Procedia Computer Science*, vol. 1, no. 1, pp. 573–581, 2012.
- [24] K. Ji, R. Sun, X. Li, and W. Shu, “Improving matrix approximation for recommendation via a clustering-based reconstructive method,” *Neurocomputing*, vol. 173, pp. 912–920, 2015.
- [25] S. Wang, C. Li, K. Zhao, and H. Chen, “Context-aware recommendations with random partition factorization machines,” *Data Science and Engineering*, vol. 2, no. 2, pp. 125–135, 2017.
- [26] S. Aarathi and M. S. Sampath, “A heat diffusion method for mining web graphs for recommendations using recommendation algorithm,” *International Journal of Engineering Research & Technology*, vol. 2, no. 3, 2013.
- [27] P.-P. Wang, P.-y. Liu, and R. Wang, “Recommendation algorithm fusing mass diffusion heat conduction and time effect,” *Journal of Chinese Computer Systems*, 2017.
- [28] Z. Shuang, B. Sheng, and S. Gengxin, “Mass diffusion recommendation algorithm based on multi-subnet composited complex network model,” in *Lecture Notes in Computer Science, Artificial Intelligence and Security*, pp. 263–274, Springer, Cham, Switzerland, 2019.
- [29] S. Aarathi and M. S. Sampath, “A heat diffusion method for mining web graphs for recommendations using recommendation algorithm,” *International Journal of Engineering Research & Technology*, vol. 2, no. 3, 2013.
- [30] Z. Haiping, S. Shigen, H. Longjun, and Z. Hongbo, “Enhanced hybrid recommendation algorithm based on mass diffusion and heat conduction processes,” *Application Research of Computers*, 2018.

## Research Article

# Numerical Investigation on the Dynamic Characteristics of an Adjustable Power Turbine Used in Environmental Control System

Qihang Lu <sup>1</sup>, Qi Zhang,<sup>2</sup> and Dalin Zhang <sup>1</sup>

<sup>1</sup>College of Aerospace Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China

<sup>2</sup>Aviation Industry Corporation of China, Nanjing Engineering Institute of Aircraft Systems, Nanjing, China

Correspondence should be addressed to Dalin Zhang; zhangdalin@nuaa.edu.cn

Received 6 July 2020; Revised 27 August 2020; Accepted 11 September 2020; Published 28 September 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Qihang Lu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, the environmental control system of aircraft driven by a power turbine is further analyzed. Through the numerical simulation of the change of the bleed state under different flight conditions and the change of the flow field under different nozzle opening, the simulation results are verified by the experimental results, and the specific change rules of the power turbine output torque, power, and bleed flow are obtained. It is analyzed quantitatively that adjusting the adjustable nozzle ring can keep the output power stable, widen the flight envelope, and improve the stability of the environmental control system.

## 1. Introduction

The main purpose of the aircraft environmental control system is to provide comfortable environmental conditions for airborne personnel and electronic equipment. By controlling the pressure, humidity, and temperature in the cabin, the environmental control system makes the system environment in the comfortable range of the human body and ensures that the electronic equipment always works within the allowable temperature range. According to the research, the traditional environmental control system needs a lot of nonpropulsion energy from the engine, and the high-temperature and high-pressure gas produced by the engine compressor is used to drive the environmental control system. At present, in order to use energy more efficiently and economically, further system optimization of the environmental control system has become a research hotspot [1, 2]. At present, there are many methods proposed by researchers but mainly focus on the following two directions:

(1) Continue to drive the environmental control system by bleed air from the engine compressor. Improve the performance of components or use power turbine to replace the traditional air entrainment method, so as to reduce energy waste [3].

(2) By building multielectric or all-electric aircraft, use as much electricity as possible to drive equipment on the aircraft. According to the thermal load and pressure requirements of the cabin, the environmental control system adjusts the motor to control the compressor pressure ratio to obtain the appropriate air supply temperature and pressure. The environment control system with all-electric drive works more stably and the comfort of passengers is improved significantly. At the same time, the all-electric environmental control system does not need to draw air from the engine compressor, so it will not affect the engine performance [4].

In the flight energy-saving design, the overall system optimization is very important because the single subsystem optimization cannot get the optimal solution, so in 2008, the United States launched the integrated aircraft energy technology project [5]. In the INVENT program, the concept of F-35 adaptive power supply and thermal management system (APTMS) was proposed. APTMS improves emissions by using exhaust gas from power turbines and engine compressors. The gas from these engine compressors is directly discharged after expanding and working in the power turbine. The environmental control system driven by

the power turbine was first verified on the F-15 fighter [6, 7]. The original F-15 environmental control system was an open-air circulation refrigeration system, driven by the bleed air from the engine compressor. The main difference between the F-15 and F-35 environmental control systems is the treatment of the gas leaving the power turbine [8]. The gas discharged from the power turbine of F-15 needs to enter the engine room for refrigeration; because of the requirements of the cabin pressure system, the expansion ratio of the power turbine will be relatively small. The power turbine of F-35 directly releases the gas into the external atmosphere and obtains the maximum expansion ratio and power. When the engine compressor bleed parameters are identical, the power turbine of F-35 will output more work.

Adjustable nozzle ring turbines are currently mainly used in variable geometry turbochargers for automobiles. The variable geometry supercharger can continuously adjust the effective flow cross-sectional area of the turbine inlet according to the engine operating conditions, change the turbine inlet airflow parameters, and then change the supercharging pressure, so as to achieve a good match between the supercharger and the engine and also improve the engine transient responsiveness, reducing transient emissions. Compared with the operating characteristics of the variable geometry turbocharger and the traditional supercharger, the former is more efficient and economical in the entire flow range. Improving the adaptability of the turbine under variable operating conditions is a popular direction for the development of turbine technology. However, at present, there are few studies on power turbines using variable nozzle replacement used in aircraft environmental control systems [5]. This paper further studies the dynamic performance of the power turbine in the aircraft environmental control system after the variable nozzle ring is used and the change rule of the variable nozzle ring when the bleed air pressure and temperature change.

## 2. System Analysis

The environmental control system (Figure 1) in which the power turbine is located is an improved three-wheel high-pressure water removal system (Figure 2) [9]. The environmental control system directly introduces the engine bleed gas into the power turbine, which is directly discharged into the ambient atmosphere after the gas expansion work, and the power turbine directly outputs the shaft work to the turbine compressor assembly [10]. Fresh gas is pumped directly from the surrounding atmosphere by the compressor and supplied to the cabin.

In the traditional environmental control system, after the air is discharged from the engine, the absolute pressure regulator is needed to control the pressure, and the temperature of the future gas passing through the precooler is controlled within a certain range [11]. This process wastes a lot of energy. The environmental control system driven by the power turbine exhausts air directly from the engine, thus eliminating the need for precooler and absolute pressure regulator, which not only realizes the full utilization of energy but also reduces the weight of the system [12].

The bleed pressure and temperature of the engine will change with the flight state. When the bleed state changes, the performance of the power turbine can be controlled by adjusting the opening of the variable nozzle ring, so as to ensure that the output power and torque of the power turbine remain unchanged, so as to isolate the influence of the change of engine operating state on the environmental control system. More importantly, in low altitude or ground operation, because of the low bleed pressure, the system can be operated normally in a larger flight envelope by adjusting the nozzle to obtain a larger starting torque.

This paper calculates the torque, mass flow, and power of the power turbine in the flight curve based on the engine operating parameters of a certain type of flight. When the flight state changes, the torque, mass flow rate, and power change and the adjustment of nozzle opening are calculated. The flight profile used in the calculation is shown in Figure 3.

## 3. Calculation Model

The blades of the nozzle ring are usually modeled using aerodynamic blades, linear blades, and wing blades. Among them, the pneumatic blade has the largest flow range and the highest aerodynamic efficiency, but the manufacturing process is complex and the blade width is small, which is not conducive to the installation of the rotating shaft [13]. In this paper, NACA0065 symmetric airfoil blades are used in the nozzle ring of the power turbine. The basic parameter model of the power turbine is shown in Table 1, and the three-dimensional model of the turbine and nozzle ring is shown in Figures 4 and 5.

The power turbine uses gas expansion to output torque and power, which drives the load on the shaft to work together. When the turbine is started from a standstill, the torque generated by the gas impact and the friction torque of the bearing must satisfy the following relationship [14]:

$$M_t > M_l, \quad (1)$$

where  $M_t$  is the torque of the power turbine and  $M_l$  is the drag torque of the system to the turbine.

The rotation equation of the power turbine is

$$I_c \frac{d^2\theta}{dt^2} = \Sigma M_z, \quad (2)$$

where  $I_c$  is the total moment of inertia and  $\Sigma M_z$  is the total external torque received by the power turbine, including the gas generating torque and the external load and friction torque.

## 4. Simulation Parameters

*4.1. Simulation Model and Meshing.* This article uses ANSYS ICEM software to model and mesh the power turbine in three dimensions. As shown in Figure 6, the model is divided into three parts: a volute, a diffuser, and a blade, and the interface is used as an interface to connect the three models.

In FLUENT, the dynamic numerical simulation of the power turbine is carried out using the sliding grid method.

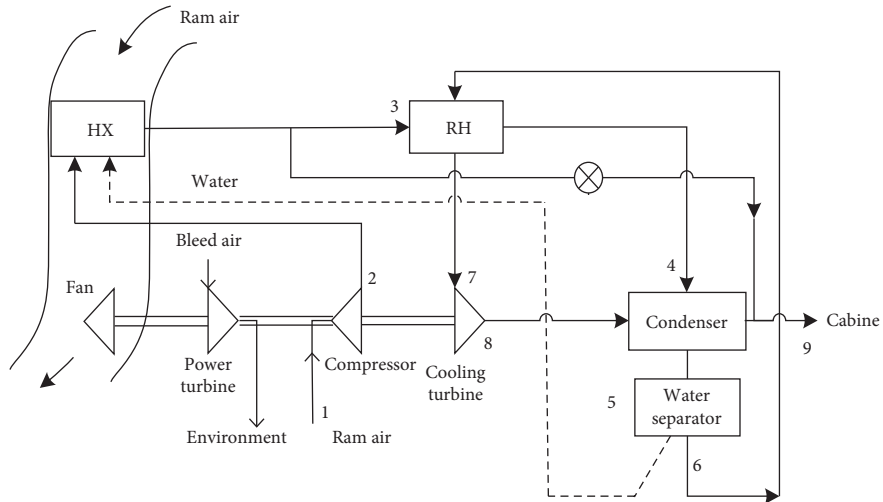


FIGURE 1: Environment control system driven by a power turbine.

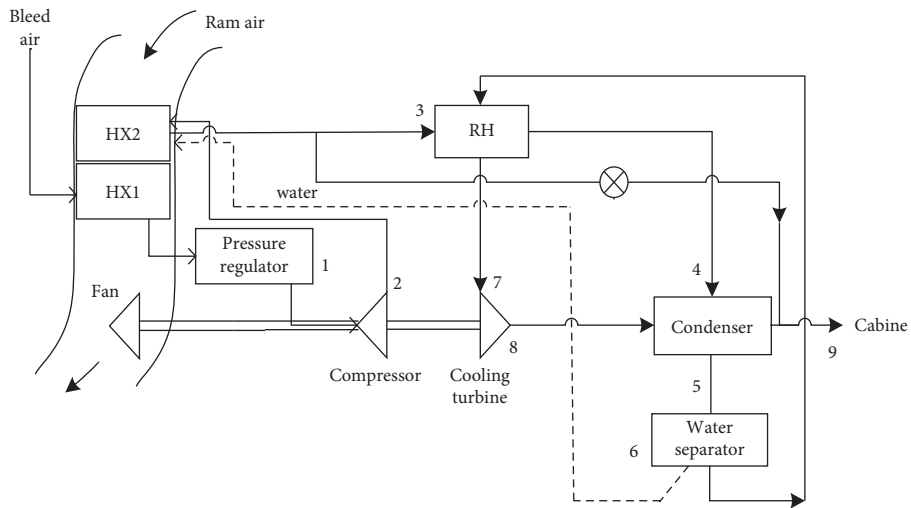


FIGURE 2: Three-wheel boost system.

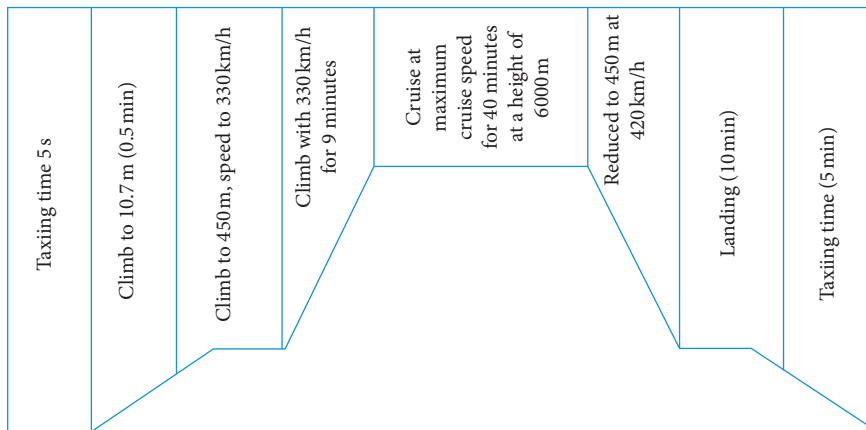


FIGURE 3: Typical mission profile.

TABLE 1: Basic parameters of the power turbine.

| Parameter                             | Value | Parameter                        | Value |
|---------------------------------------|-------|----------------------------------|-------|
| Impeller blade number $Z_i$           | 16    | Nozzle inner diameter $D_{1n}$   | 99    |
| Nozzle ring outlet blade height $I_n$ | 5.2   | Nozzle throat width $b_n$        | 4.5   |
| Impeller outlet hub diameter $D_{2h}$ | 24    | Impeller exit blade height $I_2$ | 14    |

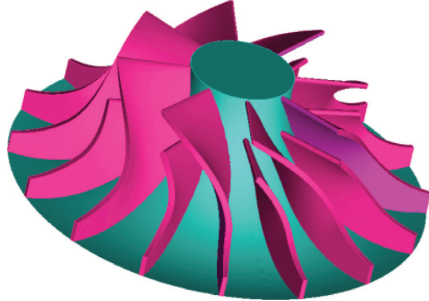


FIGURE 4: Power turbine.

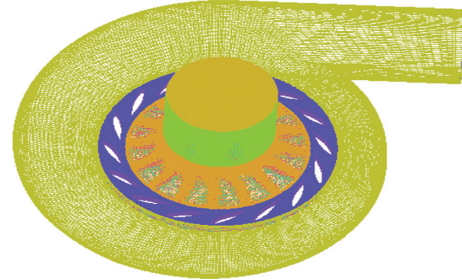


FIGURE 6: Power turbine.

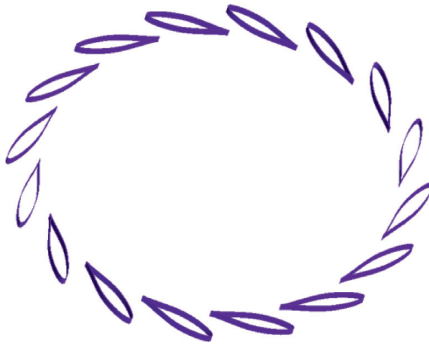


FIGURE 5: Variable area nozzle.

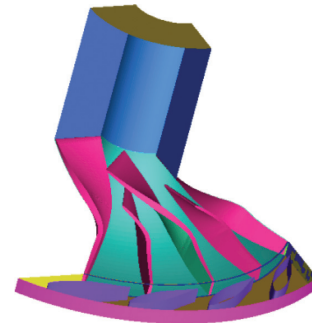


FIGURE 7: Three-dimensional modeling.

In order to obtain better simulation results, the first-order steady-state method is used for calculation, and then it is converted to second-order dynamic calculation after convergence.

The number of power turbine model grids is large, and the convergence speed is slow when using parallel computing in FLUENT. In order to efficiently obtain the performance of the power turbine with different variable nozzle ring openings, the geometric model of the power turbine must be simplified. The common method is to ignore the influence of the intake volute and use the turbine single-flow model to mesh. In this paper, the optimization calculation model uses the 1/4 periodic flow channel model of the power turbine for calculation. 1/4 nozzle ring model and impeller model are selected, and the number of grids is about 700,000 for calculation. The number of corresponding nozzle ring blades in the 1/4 flow channel model is 4 and the number of impeller blades is 4, as shown in Figure 7. Figure 8 is the mesh file inside the runner after the nozzle ring grid and the impeller grid are stitched together. When setting the boundary conditions in FLUENT, the main difference between the 1/4-channel power turbine model and the complete machine model is that the former requires a closed geometric model by adding periodic boundaries in the flow area at the left and right boundaries of the channel. The



FIGURE 8: Grid model.

boundary attribute of the nozzle ring was changed to the pressure inlet, and the definition of the remaining boundary conditions was consistent with the fluent numerical simulation of the whole machine.

**4.2. Turbulence Model.** The SA model is complete, and the turbulence scale in the model is automatically defined, and no specific flow is required to adjust the closure constant. The effect is better in the prediction of far wake and mixed laminar flow, boundary layer flow, backward step flow and

airfoil flow with small separation. In general, the SA model is more suitable for engineering applications, especially in the airfoil and impeller applications. The SA model is used for calculation in this paper:

$$\begin{aligned} \frac{\partial \tilde{\nu}}{\partial t} &= C_{b1} [1 - f_{t2}] \tilde{S} \tilde{\nu} \\ &+ \frac{1}{\sigma} \left[ \nabla \cdot ((\nu + \tilde{\nu}) \nabla \tilde{\nu}) + C_{b2} (\nabla \tilde{\nu})^2 \right] \\ &- \left( C_{w1} f_w - \frac{C_{b1}}{k^2} f_{t2} \right) \left( \frac{\tilde{\nu}}{d} \right)^2 + f_{t1} \Delta U^2. \end{aligned} \quad (3)$$

The turbulent viscosity coefficient  $\mu_T$  is calculated by the following formula:

$$\mu_T = \rho \tilde{\nu} f_{v1}, \quad (4)$$

$$f_{v1} = \frac{\chi^3}{\chi^3 + C_{v1}^3}, \quad (5)$$

$$\chi = \frac{\tilde{\nu}}{\nu}. \quad (6)$$

$\nu$  is the viscosity coefficient of molecular motion and the generated term is

$$\tilde{S} = \Omega + \frac{\tilde{\nu} f_{v2}}{\kappa^2 d^2}, \quad (7)$$

$$f_{v2} = 1 - \frac{\chi}{1 + \chi f_{v1}}, \quad (8)$$

$$\Omega = \sqrt{\left( \frac{\partial w}{\partial y} - \frac{\partial v}{\partial z} \right)^2 + \left( \frac{\partial u}{\partial z} - \frac{\partial w}{\partial x} \right)^2 + \left( \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \right)^2}, \quad (9)$$

where  $\Omega$  is the vorticity.

The function  $f_w$  is defined as

$$f_w \equiv g \left[ \frac{1 + C_{w3}^6}{g^6 + C_{w3}^6} \right]^{1/6}, \quad (10)$$

$$g = r + C_{w2} (r^6 - r), \quad (11)$$

$$r = \frac{\tilde{\nu}}{\kappa^2 d^2 \tilde{S}}. \quad (12)$$

The function  $f_{t1}$  is defined as

$$f_{t1} = C_{t1} g_t \exp \left[ -C_{t2} \frac{\omega_t^2}{\Delta U^2} (d^2 + g_t^2 d_t^2) \right], \quad (13)$$

$$g_t = \min \left( 0.1, \frac{\Delta U}{\omega_t \Delta x_i} \right), \quad (14)$$

where  $d$  is the shortest distance of the solid wall.

The function  $f_{t2}$  is defined as

$$f_{t2} = C_{t3} \exp(-C_{t4} \chi^2). \quad (15)$$

Constant values in model equations are

$$\kappa = 0.41,$$

$$C_{b1} = 0.1355,$$

$$C_{b2} = 0.622,$$

$$\sigma = \frac{2}{3},$$

$$C_{v1} = 7.1,$$

$$C_{w1} = 3.239,$$

$$C_{w2} = 0.3,$$

$$C_{w3} = 2,$$

$$C_{t1} = 1,$$

$$C_{t2} = 2,$$

$$C_{t3} = 1.1,$$

$$C_{t4} = 2. \quad (16)$$

**4.3. Parameter Settings.** Set the nozzle ring installation angle to  $12^\circ$  and the number of nozzle ring blades to 16. The change of the nozzle ring blade opening degree “+” means that the nozzle ring is closed and the flow rate decreases; the nozzle ring blade opening degree change “-” means that the nozzle ring is open and the flow rate increases. There are 7 calculation states of nozzle ring opening:  $+10^\circ + 6^\circ + 8^\circ + 3^\circ, 0^\circ, -3^\circ, -6^\circ$ , of which  $0^\circ$  is the design state, as shown in Figure 9.

The inlet pressure in FLUENT is set to 3 bar, 3.5 bar, 4 bar, 4.5 bar, and 5 bar and the outlet pressure is set to 1 bar. The starting torque received by the impeller is calculated by udf.

**4.4. Grid Independence Verification.** Table 2 can be drawn; when the number of grids is 3 million, the calculated value of the turbine efficiency is very different from the calculated value by the number of grids 4 million and 5 million. After the number of grids exceeds 4 million, the increase in the number of grids has little effect on the power turbine efficiency. It is considered that the 4 million grids have reached grid independence.

Using the same data, perform a numerical simulation on the 1/4-channel power turbine model and compare the calculation results with the numerical simulation results of the whole machine as shown in Table 3. It can be seen that the difference between the calculation results of the flow rate and efficiency between the two is very small. In the numerical simulation, the calculation result of the 1/4 flow channel can be approximately used as the numerical simulation result of the power turbine.

**4.5. Experimental Program.** Figure 10 is a traditional 3-wheel environmental control system, consisting of a turbo compressor, an electric compressor, and a radiator. The pressure

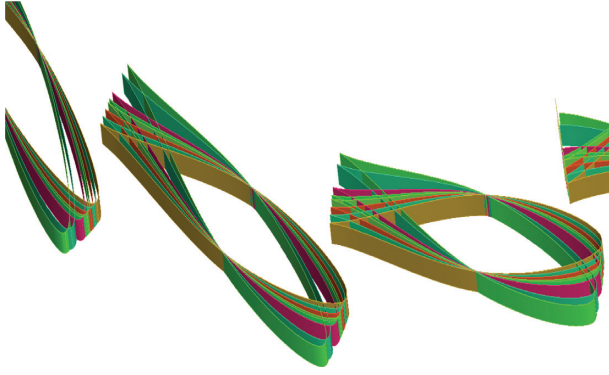


FIGURE 9: Variable nozzle ring.

TABLE 2: Different grid number numerical simulation results.

| Number of grids | Mass flow (kg/s) | Expansion ratio | Adiabatic efficiency |
|-----------------|------------------|-----------------|----------------------|
| 3000000         | 0.2453           | 3.343           | 0.8292               |
| 4000000         | 0.2452           | 3.353           | 0.8129               |
| 5000000         | 0.2452           | 3.353           | 0.8120               |

TABLE 3: Simulation results of the whole machine and 1/4 flow channel.

| Model            | Mass flow (kg/s) | Expansion ratio | Adiabatic efficiency |
|------------------|------------------|-----------------|----------------------|
| All              | 0.2452           | 3.353           | 0.8129               |
| 1/4 flow channel | 0.2472           | 3.301           | 0.8226               |

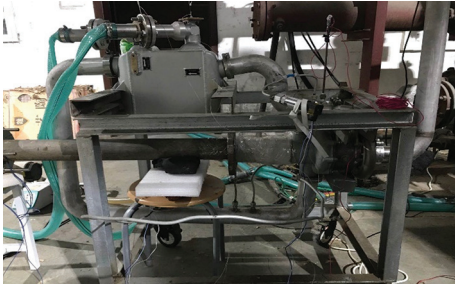


FIGURE 10: Power turbine.

sensor and temperature sensor measure the temperature and pressure of the inlet and outlet of the turbine and compressor and calculate the efficiency of its various components. As shown in Figures 10–13, the high-temperature and high-pressure gas obtained by simulating the engine bleed air is simulated by an electric compressor, and the inlet and outlet temperatures and pressures of the power turbine are measured using pressure sensors and thermocouples, respectively, to calculate the efficiency of the power turbine. The experimental value is compared with the simulation value to verify the rationality of the simulation result [15, 16]. The accuracy of the absolute pressure sensor is 2%, the accuracy of the thermocouple is  $\pm 0.5^\circ\text{C}$ , the turbine inlet pressure is usually above 200 kPa, the temperature is  $30^\circ\text{C}$ – $200^\circ\text{C}$ , the outlet is directly connected to the atmosphere, the pressure is about

100 kPa, and the temperature drop is usually  $40^\circ\text{C}$  above. The accuracy of the experimental instrument is sufficient to meet the experimental requirements.

## 5. Results and Discussion

**5.1. Experimental Results.** It can be seen from Table 4 that the simulation value is slightly higher than the experimental value by about 5%, which may be because the simulation model does not consider the efficiency reduction caused by the heat conduction of the air introduction pipe and the shell itself. The simulation value and the experimental value are basically consistent, so the aerodynamic force should be basically consistent with the actual situation.

### 5.2. Simulation Results

**5.2.1. Flow Field Analysis.** Figures 14–17 are the static pressure distribution of the gas at the nozzle ring blade height of 50% when the nozzle ring blade openings are  $+3^\circ$ ,  $0^\circ$ ,  $-3^\circ$ , and  $-6^\circ$ , respectively. It can be seen from the figure that, in the case of the four middle openings, the low-pressure area obviously appears at the position of the maximum curvature of the suction surface of the blade, and the position of the maximum load of the blade does not change. During the change of the nozzle ring opening from  $-6^\circ$  to  $+3^\circ$ , the pressure value in the center of the low-pressure area is continuously decreasing, which also causes the nozzle ring blade outlet pressure value to continuously decrease, which is consistent with the calculated nozzle ring outlet static pressure data match.

The total pressure cloud at the nozzle ring blade height of 50% is shown in Figures 18–21. It can be seen from the figure that the total pressure loss at the nozzle ring outlet is significantly larger at  $+3^\circ$ , including  $0^\circ$ ,  $-3^\circ$ , and  $-6^\circ$ . In this case, the total pressure loss is not obvious, and the change of the airflow angle of the nozzle ring blade outlet causes the loss of turbine efficiency.

The Mach number at the nozzle ring blade height of 50% is shown in Figures 22–25. The nozzle ring flow channel is a convergent flow channel. At  $+3^\circ$ ,  $0^\circ$ , and  $-3^\circ$ , the suction surface curvature is at the maximum position at the nozzle ring blade height of 50% (low-pressure area). There is a local supersonic area, which is not present at  $-6^\circ$ . The presence of a local supersonic zone will affect the uniformity of the flow in the flow field and have a certain effect on turbine efficiency.

When the nozzle ring blade opening is  $+3^\circ$ ,  $0^\circ$ ,  $-3^\circ$ , and  $-6^\circ$ , the impeller blade efficiency distribution is 0.897, 0.900, 0.894, and 0.878, which is consistent with the change trend of turbine efficiency. The static pressure cloud diagram of the long blade of the impeller is shown in Figures 26–29. It can be seen from the figure that there is a significant reverse pressure gradient area at the outlet of the impeller flow channel. As the nozzle ring blades rotate from positive opening to negative opening, the reverse pressure gradient at the outlet of the impeller flow path becomes weaker, and the pressure distribution is more uniform. In the process of changing the opening of the nozzle ring, the decrease in the efficiency of the impeller is much greater than that of the nozzle ring. The main reason is that the opening angle of the nozzle ring blades causes the inlet angle of



TABLE 4: The adiabatic efficiency of the turbine in different states.

| Inlet static pressure (kPa) | Inlet static temperature (K) | Expansion ratio | Adiabatic efficiency |                  | Error (%) |
|-----------------------------|------------------------------|-----------------|----------------------|------------------|-----------|
|                             |                              |                 | Experimental value   | Simulation value |           |
| 300                         | 357                          | 4.69            | 0.72                 | 0.75             | 4.2       |
| 300                         | 312                          | 5.52            | 0.75                 | 0.79             | 5.3       |
| 410                         | 315                          | 5.27            | 0.73                 | 0.75             | 2.7       |
| 520                         | 315                          | 6.36            | 0.81                 | 0.84             | 3.7       |



FIGURE 11: Electric compressor.



FIGURE 12: Absolute pressure sensor.



FIGURE 13: Thermocouple.

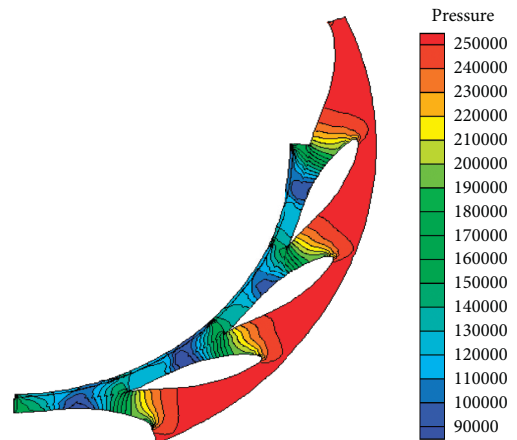


FIGURE 14: Opening +3°.

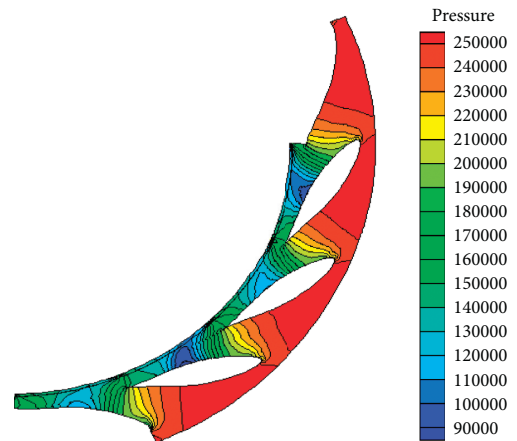


FIGURE 15: Opening 0°.

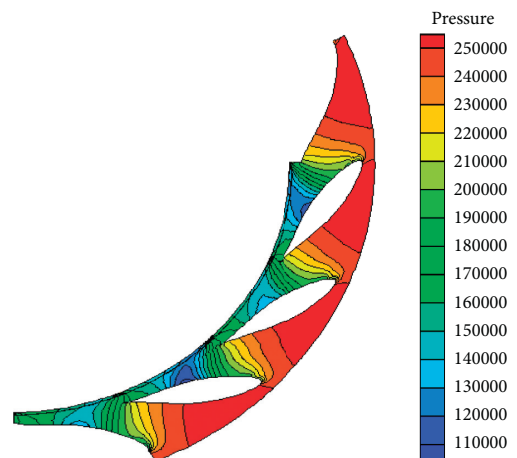
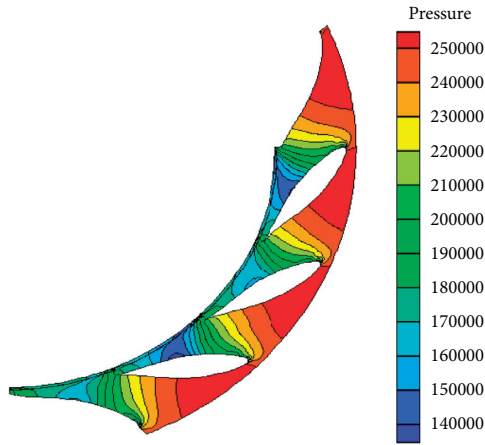
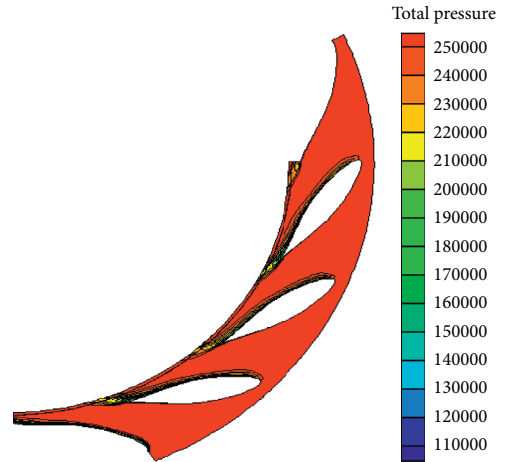
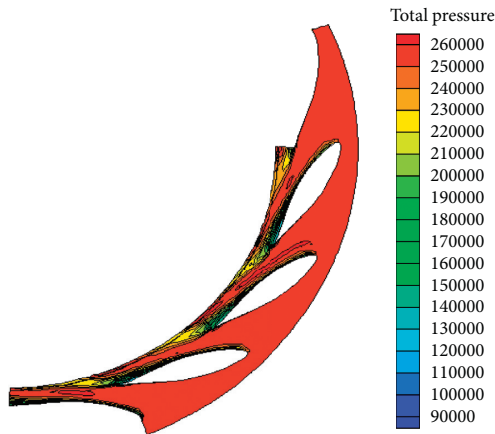
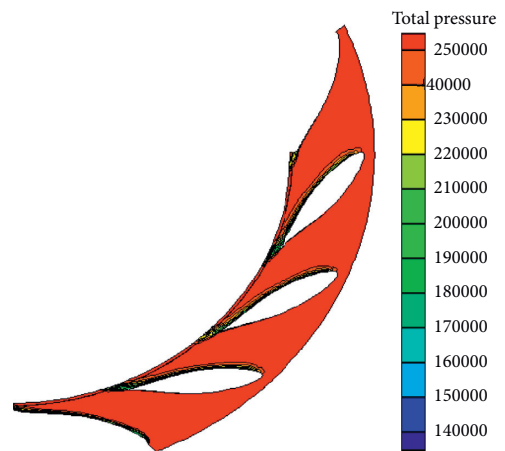
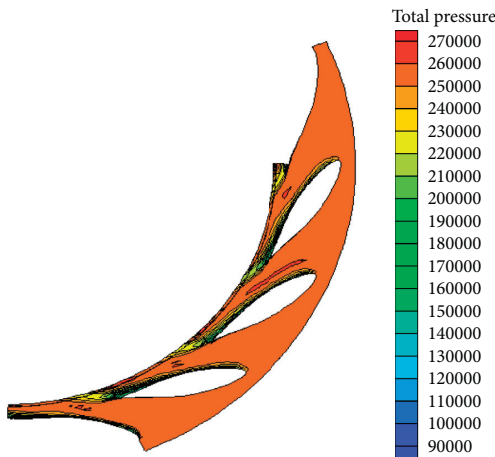
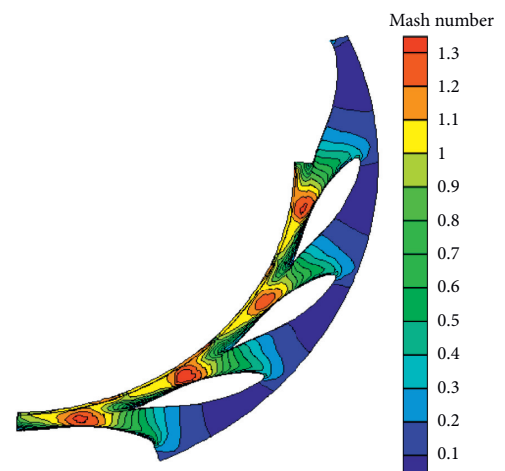


FIGURE 16: Opening -3°.

FIGURE 17: Opening  $-6^\circ$ .FIGURE 20: Opening  $-3^\circ$ .FIGURE 18: Opening  $+3^\circ$ .FIGURE 21: Opening  $-6^\circ$ .FIGURE 19: Opening  $0^\circ$ .FIGURE 22: Opening  $+3^\circ$ .

the airflow into the impeller to change significantly and because the nozzle ring blades can rotate, the gap between the nozzle ring and the impeller blades is slightly larger than that of the fixed nozzle ring turbine. The occurrence of divergence is also one of the important reasons for the decrease in the efficiency of the impeller.

**5.2.2. Torque.** As shown in Figures 30 and 31, when the inlet pressure of the power turbine is low, a larger starting torque can be obtained by opening the opening of the nozzle ring, thereby realizing the purpose of expanding

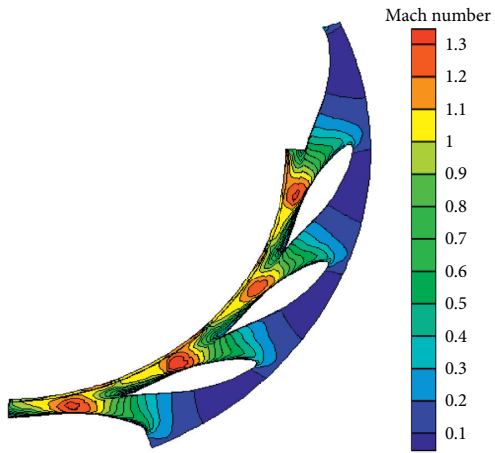


FIGURE 23: Opening  $0^\circ$ .

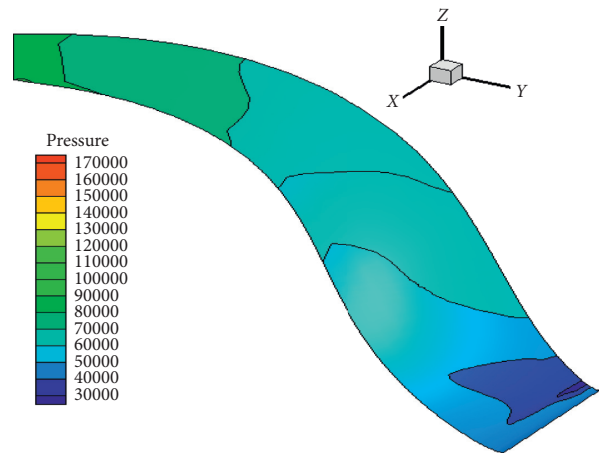


FIGURE 26: Opening  $+3^\circ$ .

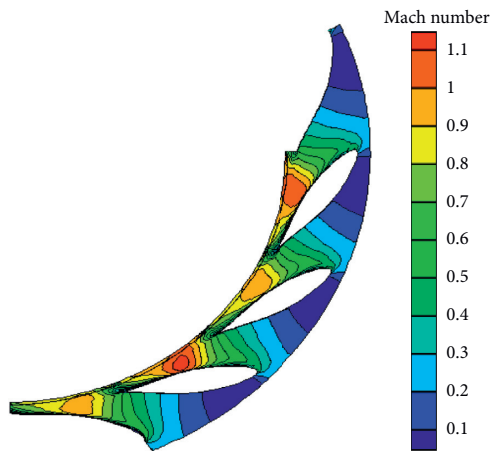


FIGURE 24: Opening  $-3^\circ$ .

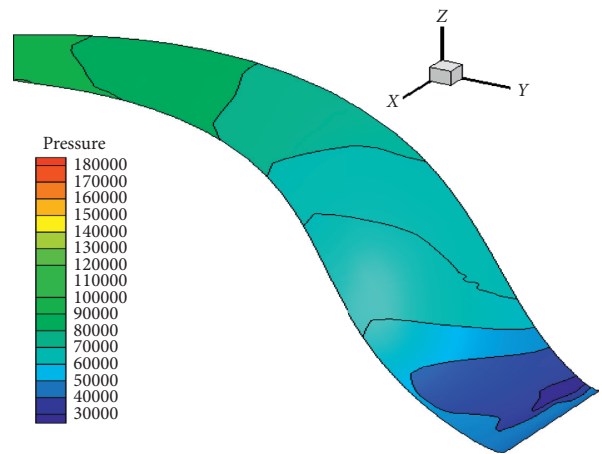


FIGURE 27: Opening  $0^\circ$ .

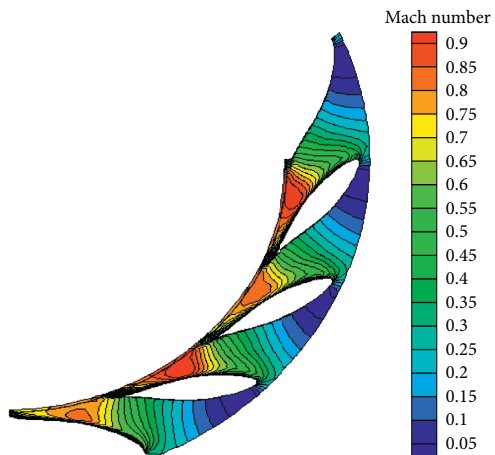


FIGURE 25: Opening  $-6^\circ$ .

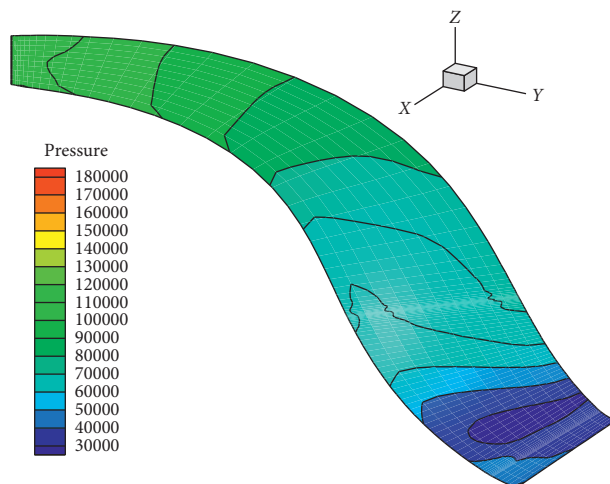


FIGURE 28: Opening  $-3^\circ$ .

the operating envelope of the system. Increasing the opening of the nozzle ring expands the system flow rate at the same inlet pressure, thereby obtaining more output torque. When the inlet pressure of the power turbine is too high, the opening of the nozzle ring can be reduced, and the system flow can be reduced, so that the torque

output by the power turbine is unchanged, and the smooth operation of the environmental control system is ensured.

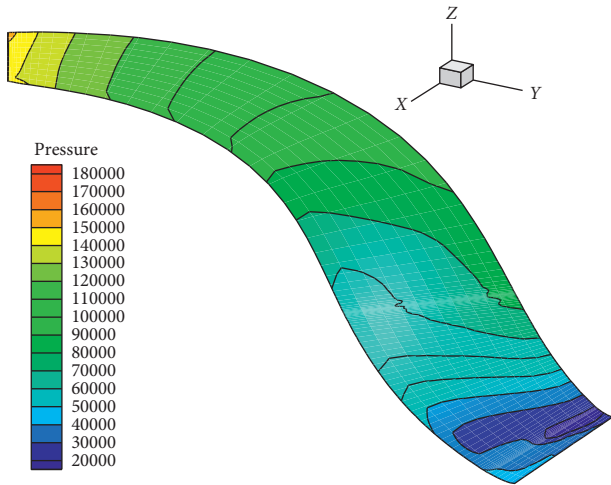


FIGURE 29: Opening  $-6^\circ$ .

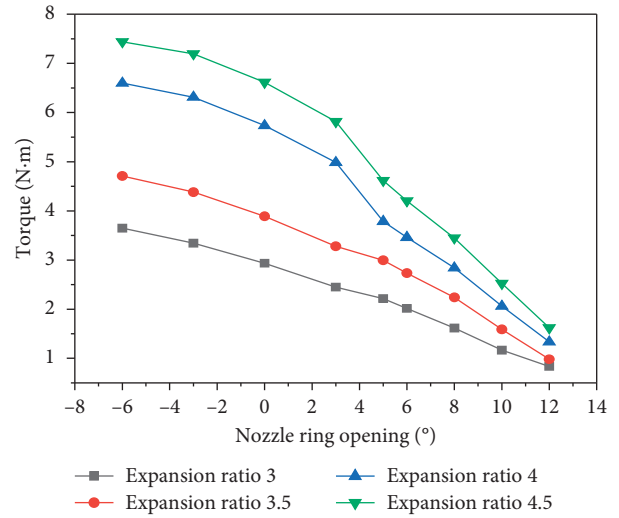


FIGURE 32: Relationship between opening and torque.

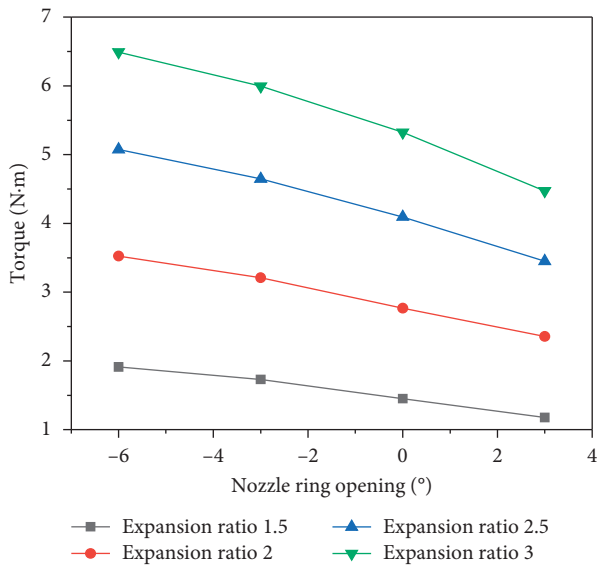


FIGURE 30: Relationship between opening and torque.

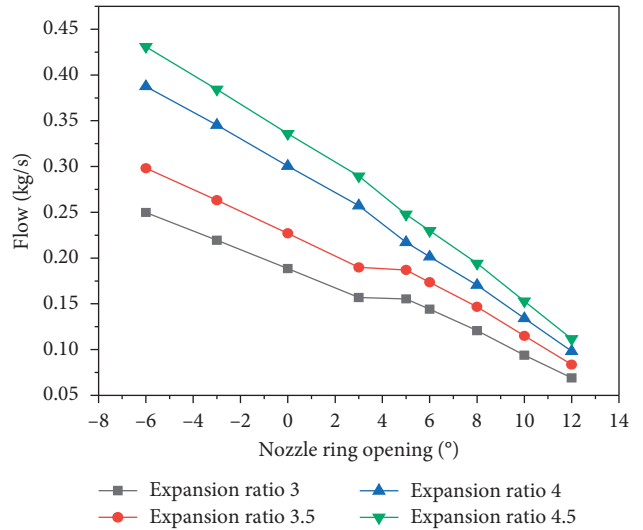


FIGURE 33: Relationship between opening and flow.

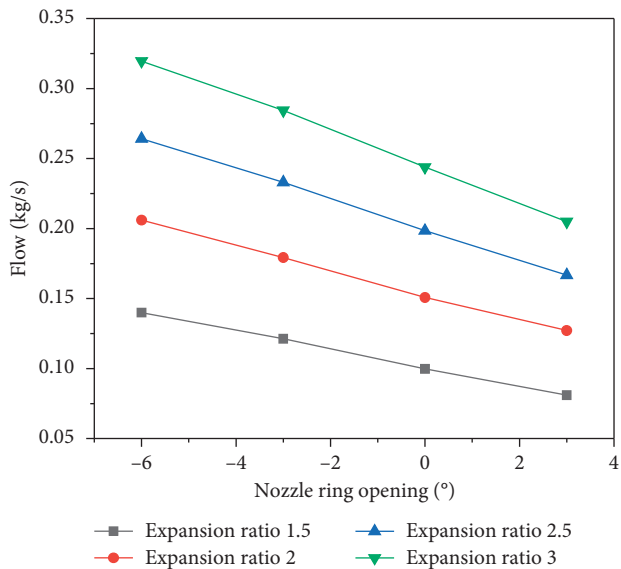


FIGURE 31: Relationship between opening and flow.

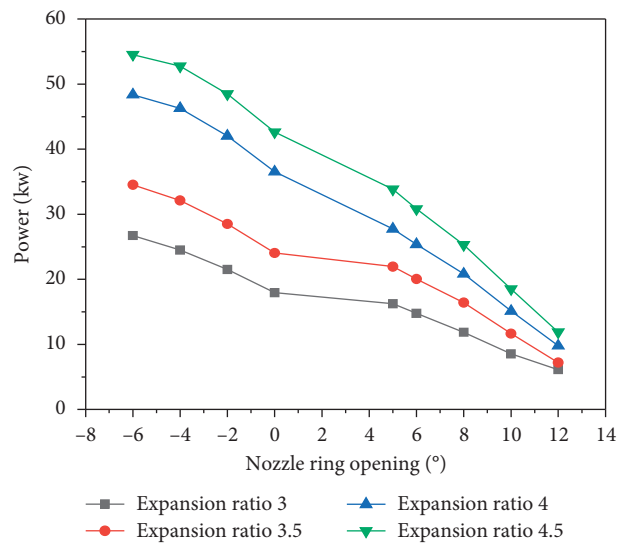


FIGURE 34: Relationship between opening and power.

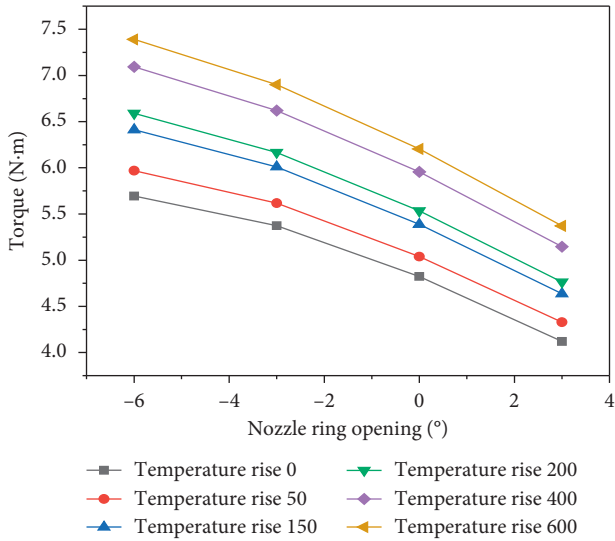


FIGURE 35: Relationship between opening and torque.

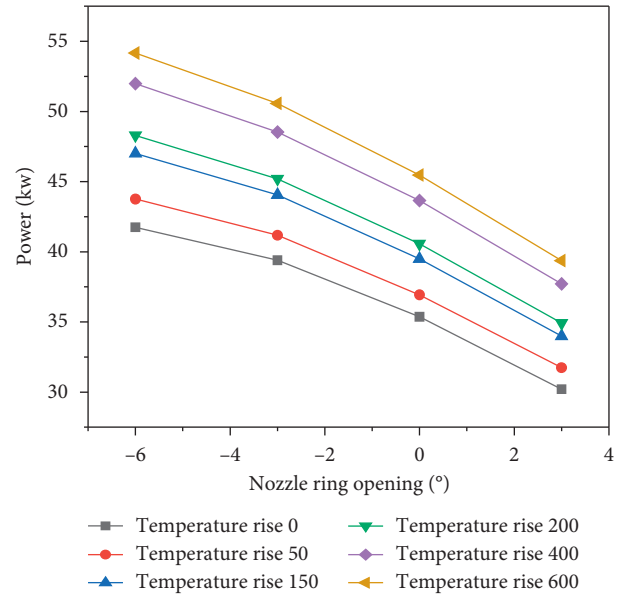


FIGURE 37: Relationship between opening and power.

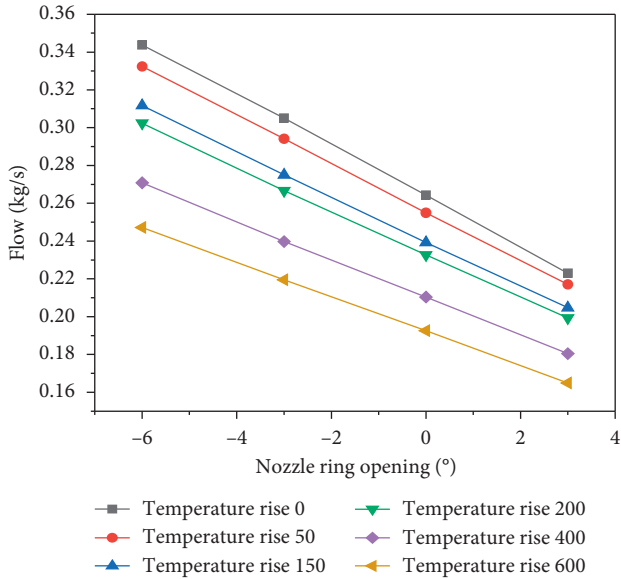


FIGURE 36: Relationship between opening and flow.

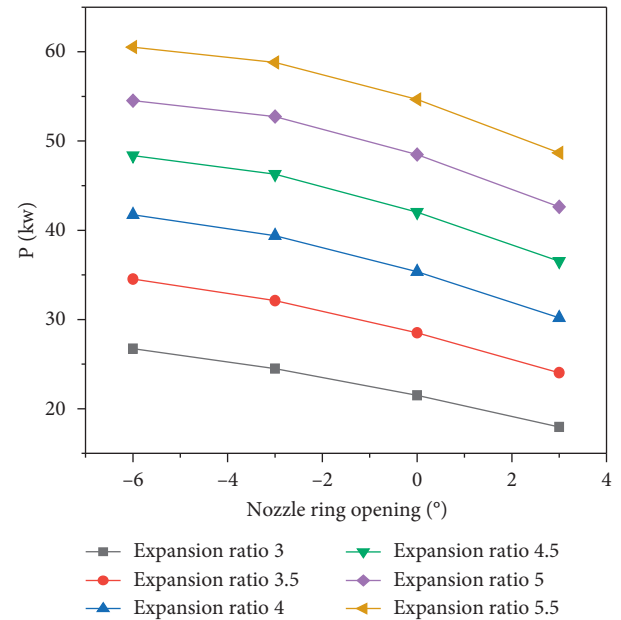


FIGURE 38: Relationship between opening and load.

5.2.3. *Constant Load and Inlet Pressure Change.* As shown in Figures 32–34, when the flight status of the aircraft changes, the bleed air pressure of the engine will change accordingly. If the opening of the nozzle ring is kept unchanged, the power and torque output by the power turbine will also change accordingly. In order to maintain the stable operation of the environmental control system, the output power and torque of the power turbine can be adjusted by increasing or decreasing the opening of the variable nozzle ring.

5.2.4. *Constant Load Change and the Inlet Temperature Changes.* As shown in Figures 35–37, when the inlet temperature of the power turbine changes, adjusting the opening of the variable nozzle can reduce the output power to match the load, thereby reducing the impact of the change in the engine bleed state on the environmental control system.

5.2.5. *Load Change and the Inlet Constant.* As shown in Figure 38, when the nozzle ring opening is changed from +0 to +6, the power can be increased by 25%, 21%, 18.1%, 15%, 12.4%, and 12.9%. When the nozzle ring opening is changed from +0 to -3, the power can be reduced by 16.5%, 15.7%, 15%, 13%, 12.1%, and 10.9%.

## 6. Conclusion

The power turbine with an adjustable nozzle ring structure can control the amount of engine bleed air according to the power required by the environmental control system. According to the flight profile, the calculated air intake of the

power turbine is significantly smaller than that of the existing aircraft environmental control system. In addition, the environmental control system can be started at a lower bleed pressure, which expands the operating envelope of the system. Therefore, a power turbine with an adjustable nozzle ring structure has the following characteristics. (1) The air circulation system driven by a power turbine with an adjustable nozzle ring structure can freely adjust the amount of bleed air according to the load, reduce fuel compensation, and realize the full use of energy. (2) The opening of the turbine nozzle can be increased or decreased to reduce the starting pressure of the power turbine and expand the range of use of the environmental control system. (3) And through the numerical simulation of the power turbine, it can be concluded that by adjusting the variable nozzle ring blades, the power turbine can provide sufficient work for the compressor throughout the flight. And when the temperature and pressure of the bleed air change, the opening can be adjusted to make the environmental control system run smoothly. In this way, the impact of changes in the engine bleed air parameters on the environmental control system is isolated.

In this paper, the gap between the nozzle ring blade and the turbine casing and the rotation axis of the nozzle ring blade are ignored in the calculation simulation. These gaps have a significant impact on turbine performance, are an important part of turbine energy loss, and also have an impact on torque. All should be considered. It is also possible to conduct research on the variable nozzle ring control actuator of the power turbine, as well as the control law of the variable nozzle ring and how to match the power turbine and the ring control system better and faster. In this paper, the research on the air circulation system driven by a power turbine with a variable nozzle ring is in its infancy, and there are still many problems to be studied in the future.

### Data Availability

The turbine's data used to support the findings of this study are included within the article. The environmental control system data used to support the findings of this study are available from the corresponding author upon request.

### Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

### Acknowledgments

This work was supported by the corresponding author Dalin Zhang.

### References

- [1] X. Su, F. Xu, and H. Ang, "Present status and future of aircraft environment control system," *Aeronautical Manufacturing Technology*, no. 10, pp. 40–46, 2002.
- [2] X. Yuan, "Developing trend discussion of control systems of high performance military aircraft," *Acta Aeronautica et Astronautica Sinica*, no. S1, pp. 2–4, 1999.
- [3] Q. Lu, D. Zhang, and S. Xiao, "Performance analysis of the power turbine in a environmental control system," in *Proceedings of the CSAA/IET International Conference on Aircraft Utility Systems (AUS 2018)*, Guiyang, China, June 2018.
- [4] R. A. Roberts and D. D. Decker, "Control architecture study focused on energy savings of an aircraft thermal management system," *Journal of Dynamic Systems, Measurement, and Control*, vol. 136, no. 4, Article ID 041003, 2014.
- [5] P. Abolmoali, J. A. Parrilla, and A. Hamed, "Integrated aircraft thermal management & power generation: Reconfiguration of a closed loop air cycle system as a Brayton cycle gas generator to support auxiliary electric power generation," SAE, Warrendale, PA, USA, Technical Paper 2014-01-2192, 2014.
- [6] S. Rajoo and R. Martinez-Botas, "Variable geometry mixed flow turbine for turbochargers: An experimental study," *International Journal of Fluid Machinery and Systems*, vol. 1, no. 1, pp. 155–168, 2008.
- [7] A. Simpson, S. Spence, and J. Watterson, "Numerical and experimental study of the performance effects of varying vaneless space and vane solidity in radial inflow turbine stators," in *Proceedings of the American Society of Mechanical Engineers (ASME)*, pp. 1495–1505, Washington, DC, USA, 2008.
- [8] D. Matullch, "High-temperature bootstrap compared with F15 growth air cycle air conditioning system," SAE, Warrendale, PA, USA, Technical Paper 891436, 1989.
- [9] F. Couaillac, "Environmental control systems for the all-electric aircraft," M.S. thesis, Cranfield University, Cranfield, UK, 2007.
- [10] S. Yu and E. Ganev, "Next generation power and thermal management system," SAE, Warrendale, PA, USA, Technical Paper 2008-01-2934, 2008.
- [11] S. Fu, Y. Zheng, K. Kan et al., "Numerical simulation and experimental study of transient characteristics in an axial flow pump during start-up," *Renewable Energy*, p. 146, 2019.
- [12] C. Liu, "Proceedings of the institution of mechanical engineers, Part I," *Journal of Systems and Control Engineering*, vol. 5, 2020.
- [13] D. Bender, "Integration of exergy analysis into model-based design and evaluation of aircraft environmental control systems," *Energy*, vol. 137, no. 15, pp. 739–751, 2017.
- [14] J. O. Gray and C. G. Downes, "A formal computer based design environment for aircraft engine control systems," *Computer Aided Design in Control Systems 1988*, vol. 21, no. 8, pp. 431–439, 1989.
- [15] J. Eichler, "Simulation study of an aircraft's environmental control system dynamic response," *Journal of Aircraft*, vol. 12, no. 10, 2012.
- [16] S. Yang, S. Chen, X. Chen, X. Zhang, and Y. Hou, "Study on the coupling performance of a turboexpander compressor applied in cryogenic reverse Brayton air refrigerator," *Energy Conversion*, vol. 122, 2016.

## Research Article

# Diagnosis and Exercise Rehabilitation of Knee Joint Anterior Cruciate Ligament Injury Based on 3D-CT Reconstruction

**Shunchao Zhang** 

*School of Sports, Pingdingshan University, Pingdingshan, Henan 467000, China*

Correspondence should be addressed to Shunchao Zhang; 3045@pdsu.edu.cn

Received 21 April 2020; Revised 23 July 2020; Accepted 3 August 2020; Published 28 September 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Shunchao Zhang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The joint capsule of the knee joint is attached to the edges of various articular surfaces and is thin and loose. Therefore, ligament reinforcement is needed to protect the knee joint and increase the stability of the joint. It plays a vital role in human activities. In this paper, a 3D-CT three-dimensional reconstruction method is used to reconstruct the ACL natural femoral imprint and double-bone tract. The relative positional relationship between the two center points is compared, and the law is summarized to guide the improvement of ACL anatomic double-beam reconstruction under arthroscopy. The 3D reconstruction results suggest that the bone layer in the anterior medial portion is the thickest, forming a peak, and the thickness of the bone layer in the posterior medial portion gradually decreases in a stepwise manner. The entire bone tissue in the anterior medial portion and posterior medial portion is integrated into one body. The tissues are connected as a whole, and the thickness is relatively uniform. The two parts of the bone tissues are not connected. The CF tissue was inserted into the bone tissue in a zigzag pattern. The changes of CF tissues in the anterior medial and posteromedial CF tissues were similar, and they were distributed stepwise from the inside to the outside. According to the bone and CF spatial structure and changing rules, ACL is divided into medial and lateral beams. According to this study, it can be summarized that (1) 3D reconstruction can clearly reconstruct the natural footprint of ACL femoral stops and postoperative osseous position and (2) 3D reconstruction can be used to evaluate the position of osseous postoperative ACL anatomic double-beam reconstruction. Arthroscopy double-beam reconstruction of ACL is instructive.

## 1. Introduction

With the increasing popularity of mass movements and the increasing incidence of traffic accidents, anterior cruciate ligament (ACL) injury is becoming more and more common, and research on it has received more and more attention from the academic community [1–3]. ACL is the core ligament that guides the movement of the knee joint. Its main function is to restrict the forward movement of the tibia, but it also plays a certain role in controlling the rotation of the knee joint, and the stress situation is complex [4]. The key part of its mechanical conduction, the tibia stop, has become the focus of biomechanical research of ligaments. Foreign scholars have reported that in animal experiments, the position of the ACL tibial plateau on the tibial plateau was moved forward and backward, and the tibial plateau and knee stability were significantly different,

suggesting the position of the tibial plateau and its restrictions [5]. The ability of the tibia to move forward and maintain the stability of the knee joint is closely related, but there is still no research report on the clear correspondence between the position of the ACL tibial stop and the ACL control force [6–9]. The connection point between the ligament and the bone is not a simple ligament bone connection, but a complex structure. The literature calls it a “stop point organ,” which can be divided into fibrocartilage stops (direct stops) and fiber stops (indirect stops). The ACL tibia stop is a typical fibrocartilage stop, including four regions, namely, fibrous tissue, uncalcified fibrocartilage (UF), calcified fibrocartilage (CF), and bone tissue. The structure mainly plays a role of stress buffering [10–12]. The domestic and foreign literature studies reporting the histological structure of ACL tibial junctions are mainly concentrated along the longitudinal axis of ACL. The purpose is

to observe the healing after surgical reconstruction. It is not a special continuous layer structure study.

The knee joint is the largest, most complex, and weight-bearing joint in the human body, and it is also the joint with the highest functional requirements of the human body. The joint capsule of the knee joint is attached to the edges of various articular surfaces and is thin and loose. Therefore, ligament reinforcement is needed to protect the knee joint and increase the stability of the joint. It plays a vital role in human activities. Ligaments are tough, band-shaped tissues connected between the femur and tibia. The knee ligaments mainly include anterior cruciate ligament (ACL), posterior cruciate ligament (PCL), medial collateral ligament (MCL), and lateral collateral ligament (LCL). The anterior cruciate ligament is the most important stabilizing factor. The anterior cruciate ligament starts from the anterior angle of the meniscus and anterior to the tibial intercondylar eminence, and ends behind the medial side of the femoral condyle. Its main functions include the following: restricting the forward movement of the tibia, preventing overextension of the knee joint, and preventing internal and external rotation of the tibia [13–15]. When the external force applied to the knee joint exceeds the anterior cruciate ligament tolerance, anterior cruciate ligament injury occurs. Anterior cruciate ligament injury has been common in traffic accidents and professional athletes in the past, but with the improvement of social living standards, sports-loving, as the number of ordinary people increases, knee joint sports injuries also increase. Among them, sports injuries such as anterior cruciate ligament rupture are more and more common in the general people. In Europe and the United States, anterior cruciate ligament injury occurs mostly in fierce sports such as basketball and football. In domestic studies, anterior cruciate ligament sports injuries are most common in basketball, football, and jumping sports. The damage of anterior cruciate ligament between men and women is also different, which is related to their sports hobbies, sports technology, and sports popularity requirements. Imaging evaluation plays a key role in the diagnosis, treatment choice, and follow-up of kidney stones. Over the years, as the incidence of urinary tract stones has increased and the understanding of disease processes has increased, imaging techniques have continued to evolve. CT urography (CTU) developed on the basis of CT is a method of injecting a contrast agent into a vein, using the secretory function of the kidney and using CT to observe the degree of filling of the contrast agent over time in the renal calamus, renal pelvis, ureter, and bladder, as shown in Figure 1. Obtaining continuous tomographic data and then three-dimensional reconstruction of the original image using a computer to obtain a three-dimensional urinary system image has become a more common inspection method in urology [16, 17]. Preoperative imaging has a crucial influence on the choice of the surgical method. However, the CTU examination results uploaded to the doctor’s workstation by the radiology department currently only have 3D reconstructed images with several perspectives. The surgeon cannot perform more operations on the 3D reconstructed model and cannot meet the needs of the surgeon. In severe cases, the

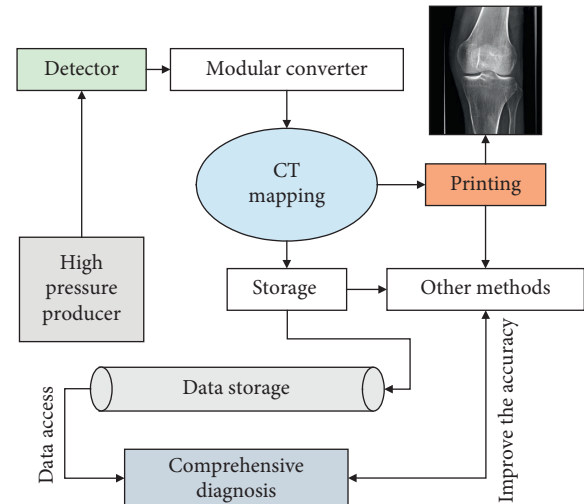


FIGURE 1: The schematic diagram of 3D-CT mapping used in the ACL detecting.

contrast agent cannot be well distributed in the pyloric renal pelvis, which may lead to poor CTU image quality, and some urinary tracts cannot be displayed. Doctors rely more on plain CT scans for preoperative planning, especially for complex situations. In more detailed preoperative planning, how to obtain better preoperative planning information is obviously the problem that the operator needs to solve first [18].

For medical images such as CT data reading and 3D reconstruction, part of the software comes from the CT machine of Siemens and other companies [19]. Usually, a machine corresponds to a software system, and individual users are generally unable to use it [20]. For external 3D reconstruction software, the current recognition is the Mimics software from the Materialise company [21]. The reconstructed model is highly consistent with the real situation. Through modular function settings, the operation is simple and easy to use, and it can directly interface with CAD, FEA, RP, and other data [22–24]. It is easy to process the model, and the software is constantly updated, and the algorithm is optimized to enhance the function. It has been widely used in orthopedics, stomatology, and other departments, mainly for individualized preoperative planning of implanted prostheses, etc., and can guide the operation in real time during the operation. 3D printing technology is an additive rapid prototyping technology. It was proposed by the Massachusetts Institute of Technology in the United States to form solids layer by layer according to a certain path [25]. With the development of materials, materials that can be used for 3D printing have also developed from traditional single solid powders such as metals and plastics to new hybrid materials such as liquids and gels. Different “inks” can be used to print different items. It has been applied in aerospace, medical, and other fields and has developed rapidly. In this paper, a 3D-CT three-dimensional reconstruction method is used to reconstruct the ACL natural femoral imprint and double-bone tract. The relative positional relationship between the two center points is



compared, and the law is summarized to guide the improvement of ACL anatomic double-beam reconstruction under arthroscopy. In addition, this study intends to perform a three-dimensional reconstruction based on the sequential tissue histological structure based on the clear tissue-stress correspondence reported in the literature, and then divide the functional bundle of the ACL, as well as the proposed positioning method for the inner center of the tibial tunnel in surgical reconstruction. It is a new method for the study of ACL function that is different from all previous reports in the literature and provides a new method and basis for future large sample research and animal experiment research.

In this paper, we firstly use the 3D-CT mapping methods to build an ideal ACL reconstruction scheme, and the innovation points can be listed as follows:

- (1) The 3D-CT three-dimensional reconstruction method is used to reconstruct the ACL natural femoral imprint and double-bone tract.
- (2) The relative positional relationship between the two center points is compared, and the law is summarized to guide the improvement of ACL anatomic double-beam reconstruction under arthroscopy.
- (3) The 3D reconstruction results suggest that the bone layer in the anterior medial portion is the thickest, forming a peak, and the thickness of the bone layer in the posterior medial portion gradually decreases in a stepwise manner. The entire bone tissue in the anterior medial portion and posterior medial portion is integrated into one body. The tissues are connected as a whole, and the thickness is relatively uniform. The two parts of the bone tissues are not connected. The CF tissue was inserted into the bone tissue in a zigzag pattern. The changes of CF tissues in the anterior medial and posteromedial CF tissues were similar, and they were distributed stepwise from the inside to the outside.
- (4) According to the bone and CF spatial structure and changing rules, ACL is divided into medial and lateral beams. According to this study, it can be summarized that (a) 3D reconstruction can clearly reconstruct the natural footprint of ACL femoral stops and postoperative osseous position and (b) 3D reconstruction can be used to evaluate the position of osseous postoperative ACL anatomic double-beam reconstruction. Arthroscopy double-beam reconstruction of ACL is instructive.

## 2. Materials and Methods

**2.1. General Information.** Among the patients undergoing three-dimensional reconstruction of the anterior cruciate ligament femoral junction treated in the affiliated hospital of Pingdingshan University from February 2018 to March 2020, 45 patients were selected, including 23 males and 22 females, aged 21 to 54 years, with an average of  $30.2 \pm 7.1$  years old (Table 1). All patients had no obvious history of

trauma to both knees, and those with knee joint swelling and pain were excluded. Patients with negative hyperextension test, lateral stress test, McMurray test, floater test, Lachman test, and knee drawer test were all negative. In addition, 45 subjects signed informed consent, and this study has been approved by the ethics committee of the hospital. In this study, a 64-row spiral CT scanner (240 mA, 120 kV) produced by GE of the United States was selected, the matrix size was  $512 \times 512$  pixels, the scan time was set to 5 s, the layer thickness was 5 mm, the window diameter was 360 mm, and the reconstruction interval layer thickness was 0.625 mm. The scan results were exported and stored in a workstation, and the 3D reconstruction software Mimics 10.01 produced by Materialise was used to construct the femur and ACL model and saved in format. Geomagic Studio12, the reverse engineering software manufactured by Geomagic, was used to analyze the SD data, implement image registration, and construct a three-dimensional solid model of cruciate ligament and attachment points of the knee joint through MRI and CT influence fusion methods. After the reconstruction of the model is completed, the femoral dead center is located according to the ACL three-dimensional model, and the ACL femoral dead center is identified and circled according to the morphological characteristics of the ACL attachment area. Geomagic software was used to record the area, long-axis angle, short distance, inner distance, back distance, short axis, and Changzhou of ACL femoral nodal points.

**2.2. Magnetic Resonance Scanning.** The magnetic resonance scanning was performed using a Siemens 3.0 T magnetic resonance instrument and an 8-channel knee joint coil. The subjects were supine with first entrance of their feet, and sagittal T1WI (repetition time (TR) 624 ms, echo time (TE) 17 ms), T2WI (TR 5069 ms, TE 113 ms), PDWI (TR 3155 ms, TE 33 ms), coronal T1WI (TR 614 ms, TE 16 ms), and PDWI (TR 3132 ms, TE 34 ms) sequence scanning were sequentially conducted. The T2-mapping imaging performed 5 echoes and 1 scanning through the SE sequence, and the sagittal scanning was conducted perpendicular to the tibial plateau. Scanning parameters were repetition time of 1500 ms, echo time of 12.4, 25.8, 39.1, 53.6, and 65.5 ms, number of layers of 24, layer spacing of 0.5 mm, field of view of  $16 \text{ cm} \times 16 \text{ cm}$ , matrix of  $368 \times 368$ , layer thickness of 5 mm, number of acquisitions of 1, flip angle of  $180^\circ$ , and acquisition time of 6.23 min. The scanning sequences were perpendicular to the sagittal plane, and the imaging data were transmitted to workstation for postprocessing to directly obtain T2-mapping pseudocolor image of the knee joint and the T2 value data measurement.

**2.3. Secondary Injury after Anterior Cruciate Ligament Rupture.** The anterior cruciate ligament of the knee joint is an important structure to maintain the stability of the knee joint's anterior and posterior rotation. If it breaks, it will seriously affect the function of the knee joint. If it is not treated in time, it will cause damage to the articular cartilage and meniscus, eventually leading to osteoarthritis happened.

TABLE 1: The general information of selected patients.

| Group  | Average age (years old) | Gender (male/female) | Disease duration (month) | Injury part |       | Patient's occupation |    |    |    |    |
|--------|-------------------------|----------------------|--------------------------|-------------|-------|----------------------|----|----|----|----|
|        |                         |                      |                          | Left        | Right | AL                   | SD | DC | FF | OO |
| ACL    | 30.2 ± 7.1              | 29/11                | 0.83 ± 0.12              | 18          | 22    | 13                   | 11 | 8  | 5  | 3  |
| Normal | 28.2 ± 2.1              | 29/11                | 0.76 ± 0.07              | 18          | 22    | 13                   | 9  | 10 | 6  | 2  |

Note: AL, athlete; SD, student; DC, dancer; FF, firefighter; OO, other occupation.

Studies have shown that the types of damage to the medial and lateral meniscus are more common after rupture of the anterior cruciate ligament, and the sites are mainly posterior horn injuries that can be sutured under the microscope. Most of the lesions need to be removed. The incidence of lateral meniscus tears is very stable and does not increase with the time after anterior cruciate ligament rupture. This suggests that anterior cruciate ligament injury may be accompanied by lateral meniscus injury. Incidence rate can rise to more than half within 1 month after anterior cruciate ligament rupture. Within 1 to 3 months after anterior cruciate ligament rupture, the patient is subjected to knee braking after trauma. Partial healing of the anterior cruciate ligament may cause the incidence of lateral meniscus tears was reduced to below 40%; the incidence of lateral meniscus tears increased to more than half within 3 months to 12 months after anterior cruciate ligament rupture. The incidence of fissures in the anterior cruciate ligament caused by lateral meniscus tears for more than 12 months remained at about 40%. There was no significant difference in each time period. The instability of the knee joint will cause damage to the medial meniscus [8]. The incidence of medial meniscus injuries within 3 months of anterior cruciate ligament rupture is about 30%; medial meniscus injuries lose 3 in the anterior cruciate ligament. The incidence rate is about 40% from a month to half a year; the incidence rate of medial meniscus injury in anterior cruciate ligament injury is more than half in half a year to one year, and the anterior medial meniscus injury occurs after cruciate ligament injury time exceeds one year. The rate can reach more than 60%, and the statistical analysis of each time period has statistical differences, as shown in Figure 2. It is suggested that the time after the cruciate ligament rupture is greater than half a year, and the probability of its medial meniscus injury will increase, indicating that after anterior cruciate ligament rupture, cruciate ligament reconstruction should be performed within 6 months to prevent secondary medial meniscus damage.

#### 2.4. Surgical Treatment of Anterior Cruciate Ligament Injury.

Anterior cruciate ligament reconstruction under knee arthroscopy is currently the preferred method for treating anterior cruciate ligament injury. Commonly used anterior cruciate ligament reconstruction includes single-beam reconstruction and double-beam reconstruction, anatomical reconstruction, and nonanatomical reconstruction [26–28]. Its current positioning method is anatomical reconstruction, including anatomical single-beam reconstruction and anatomical double-beam reconstruction. It also evolved from

the earliest nonanatomical reconstruction (nonanatomical single-beam reconstruction and nonanatomical double-beam reconstruction). The positioning of the bone canal is crucial in the process of anatomical reconstruction. First, see if you can see the broken end of anterior cruciate ligament injury to locate the femoral tract. If the broken end exists, you can use the method of central positioning of the broken end.

If the broken end does not exist, you can perform epiphyseal positioning. Resident Thallium and fasciatus are the main basis. The study found that about 80% of patients can only identify resident doctors, about 55% of patients can only identify fasciatus, and only about 50% of patients can identify both tadpoles at the same time. Anatomical studies of cadaver specimens vary. Therefore, not every patient can clearly identify the two in the knee joint. Chen Lianxu's experience is that he wants a better observation effect [29]. Arthroscopy can be accessed from the anterior internal approach and then use the hook to explore the stump of the anterior cruciate ligament to find the bulge and hard touch, as shown in Figure 3. Here, the AMB means the advanced material buffer used to improve the bone cross section. At present, most scholars insist on using single-beam reconstruction of anterior cruciate ligament. After long-term research and development, arthroscopic single-beam reconstruction of cruciate ligament has achieved satisfactory surgical results. However, due to the lack of single-beam reconstruction of anterior cruciate ligament in controlling the rotation stability of the knee joint, some scholars believe that double-beam reconstruction has obvious advantages in theory and advocate the use of double-beam reconstruction of anterior cruciate ligament [30]. However, whether the clinical efficacy of the double-beam reconstruction technique is better, the single-beam reconstruction technique has not been proved by large-scale clinical comparative trials. At present, there are mainly three kinds of grafts for anterior cruciate ligament reconstruction: allografts, autograft, and artificial ligaments. Discussion on the selection of anterior cruciate ligament grafts is mainly focused on the allograft and autograft. At present, for the selection of grafts, the comparison of the efficacy of autograft or allograft in anterior cruciate ligament reconstruction has no conclusion yet. Due to the relatively lack of reports on clinical research of synthetic grafts, and synthetic grafts have not been clinically validated due to factors such as their safety hazards and treatment effects, they are not currently the first choice by clinicians, so how to choose anterior cruciate ligament graft pair to the clinical treatment effect and clinical treatment plan has important influences [31–33]. For the graft, there are many choices for our doctors to consider,

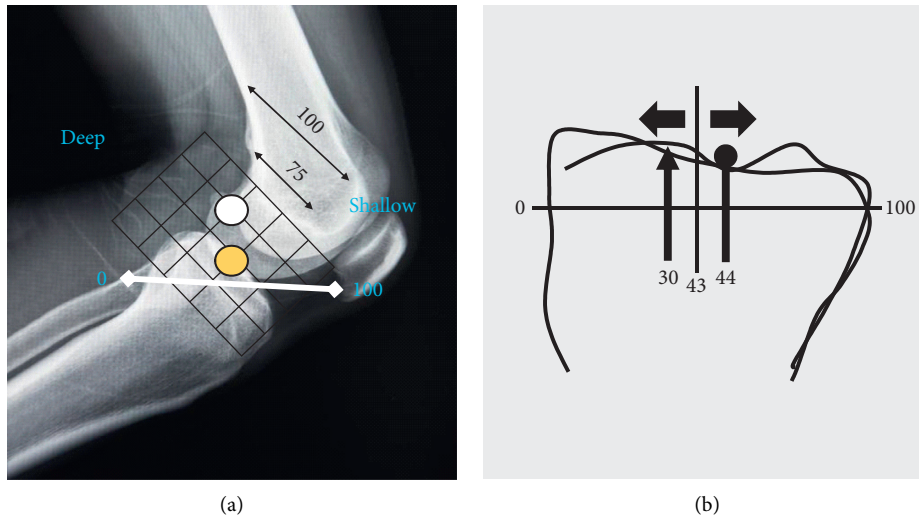


FIGURE 2: The ideal tibial bone tunnel positions for the secondary injury after anterior cruciate ligament rupture.

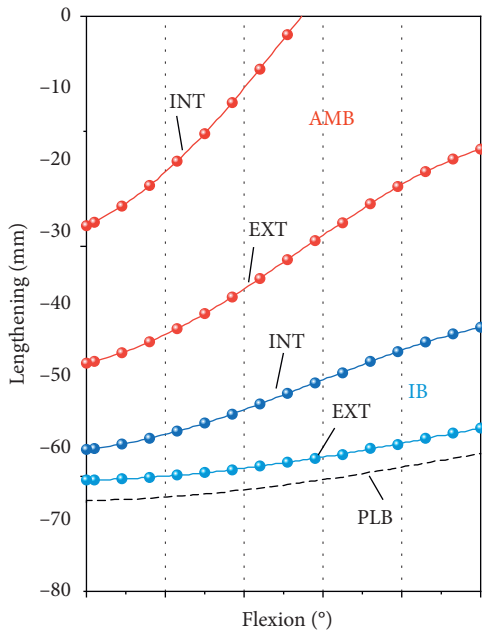


FIGURE 3: Mean length change patterns for the three ACL bundles during flexion, in internal (INT) and external (EXT) tibial rotation after surgical treatment of anterior cruciate ligament injury.

including the potential risk of infection, the hidden dangers of the need for a second surgery on the patient, and the mobility and stability of the knee joint after reconstruction. However, clinicians did not conduct too many screenings when selecting grafts due to various factors. At present, whether it is an allograft or an autograft ligament, its good clinical effect has been proven. However, the comparison of which type of graft to rebuild is preferred, which is inconclusive. For postoperative comparisons between autograft and allograft, the conclusions obtained in different studies are quite different. The main conclusions are divided into two types: one is that the choice of different grafts does not affect the function of the knee joint after patient

reconstruction, and the other is that patients who choose autograft reflect a better recovery of knee function, which may have potential infection risks and postoperative ligaments. Slow healing is the main problem for the choice of allografts, but it has been temporarily confirmed that the chance of rejection of allografts is not significantly different from that of autograft. Although autograft as a graft for reconstruction of knee ligament is still the first choice of most clinicians from the current research and clinical experience, it is important to note that the surgical effect of allograft and autograft is important. It is the surgical method chosen, how to treat the graft and the source of the graft, etc. These factors are also the main reasons affecting the effect of anterior cruciate ligament reconstruction. At the same time, allografts have a wide range of sources and are available in various sizes for clinical use. Selection, the operation during the operation is easy, and these advantages can shorten the operation time.

2.5. *Observation Indicators.* All data analysis was performed in the statistical software SPSS18.0. In this study, the count data were expressed as a percentage and tested. The measurement data were expressed as mean  $\pm$  standard deviation ( $\pm s$ ). Differences were considered statistically significant at  $P < 0.05$ .

### 3. Result Analysis

3.1. *Morphological Typing Results.* In this study, 45 patients with a total of 90 knees received 3D reconstruction analysis. The morphological classification showed that 44 knees were oval, 15 knees round, 16 knees crescent, 5 knees long, and 10 knees irregular. The oval shape is the most common, accounting for 48.89% of the total. Arthroscopy ACL reconstruction surgery is a common type of knee joint surgery. In the past, four-quadrant method, Takahashi method, and interpositional method were used to reconstruct the femoral tract, which cannot be applied in clinical promotion. With

the gradual popularization of the concept of anatomical reconstruction in recent years, due to the differences in the position, shape, and size of ACL femoral stops, it is very important to discuss how to implement effective 3D reconstruction of ACL femoral stop significance. Previous studies on ACL femoral nodal points were mostly performed on corpses. For example, Lorenz et al. used 12 corpses as samples for ACL anatomy analysis. Due to the limited number of samples and the age of the samples, the reliability of the ACL anatomy data obtained is low. The continuous development of clinical imaging technology has provided new ideas for the construction of ACL models. At present, a variety of knee joint models, such as anatomical models, mechanical models, and mathematical models, have appeared in clinical practice. Among them, the three-dimensional knee joint model based on MRI and CT technology has been widely promoted in clinic with its high degree of reduction, as shown in Figure 4. Relevant data show that the three-dimensional imaging of the knee joint can achieve multiplane and multiangle observation, thereby achieving good spatial positioning characteristics and providing reliable three-dimensional data for anatomical research. The angle is the same as the ACL femoral stop morphology plane suppression, which facilitates intraoperative osseous positioning.

In previous studies, knee functional braces were considered to be an integral part of exercise rehabilitation therapy, which can support the healing of ACL, improve the functional stability of joints, and restore normal kinematics. In different exercise rehabilitation stages, when performing exercise rehabilitation function exercises, the adjustable knee joint functional brace can adjust the knee joint activities in different ranges and has a supporting and protecting effect on the affected knee joint. In clinical exercise rehabilitation training after anterior cruciate ligament reconstruction, adjustable knee joint functional braces have been widely used. However, there are still different opinions on the effects of postoperative exercise rehabilitation training on whether or not a brace should be worn. A systematic review summarizes the related research of 28 braces. The research angle is to consider whether the braces can improve the kinematic function of the knee joint. The conclusion is that the literature does not currently recommend the use of knee function exercise rehabilitation after anterior cruciate ligament reconstruction braces, and it is believed that the use of knee functional exercise rehabilitation braces after anterior cruciate ligament reconstruction will not reduce postoperative pain and improve knee function. However, for patients after anterior cruciate ligament reconstruction, the main role of the knee functional exercise rehabilitation brace is first to support graft healing and maintain knee stability, and then to perform functional exercise rehabilitation training. In addition, some scholars believe that if the postoperative patients do not wear braces for protection, knee function cannot be recovered well during exercise rehabilitation. It is recommended that patients who undergo anterior cruciate ligament surgery need to wear braces within 1 to 3 months after surgery. In the early stage of exercise rehabilitation, the brace should be mainly

used to protect the knee joint. In the later stage of exercise rehabilitation, if there is a patient who needs to return to exercise, the choice of the brace should be mainly based on functional exercise.

CT technology covers the processes of projection data acquisition, data correction, and 3D image reconstruction, and finally obtains the 3D reconstruction image of the scanned object, and the input of 3D printing is based on the 3D volume data reconstructed by CT technology and represented by a triangular grid. As shown in the figure above, the entity is scanned by CT to obtain a multiangle projection data volume. After three-dimensional reconstruction, tomographic data are obtained. After format conversion, topography analysis and STL data export are performed; STL data are combined with 3D printing to print out samples. And we realize the structural information comparison between the original sample and the 3D printed sample.

### 3.2. Comparison of Left and Right ACL Femoral Stops.

There were no statistically significant differences in the measured data of ACL femoral dead points of the left and right knees of 45 subjects ( $P > 0.05$ ), as shown in Figure 5. Here, the bone section shows almost no difference on the left and right sides. The slight difference between 23.7% and 21.7% could be only the individual error. At present, CT imaging is used in the diagnosis of a large number of soft tissues, bone, and joint lesions. CT imaging is an ideal inspection method for bone and joint lesions. It can detect and reflect the small density differences of different tissues; MRI imaging has high spatial resolution rate, has significant advantages in imaging of small tissues or structures such as blood vessels, tendons, cartilage, and nerves, and has high diagnostic sensitivity and specific mockery for knee ligament rupture and meniscus injury. In this study, the knee joint MRI and CT data were collected, combined with the advantages of MRI imaging of soft tissue and CT for bone structure, and three-dimensional reconstruction of knee ACL and femur to obtain a three-dimensional model of knee joint, so as to fully reproduce knee ACL. Compared with the traditional ACL anatomy research method, the model constructed in this study can achieve accurate measurement of ACL anatomy data, and it is more accurate and reliable in confirming ACL femoral stops. Previous studies have shown that there is no significant difference in the morphological parameters of the left and right knees of knee specimens. For example, Dargel et al. performed ACL femoral measurement on the specimen. The length of the short axis was  $(10.31 \pm 1.45)$  mm, and the long axis was  $(17.01 \pm 1.83)$  mm. In this study, the measurement results of the knee ACL three-dimensional reconstruction model of 90 subjects showed that the left and right knees had no significant difference in parameters such as the long axis, short axis, area, and long-axis angle ( $P > 0.05$ ), which is basically consistent with the conclusions in related reports, indicating that the anatomical data obtained by 3D reconstruction of knee ACL given MRI and CT data have good reliability. Some studies have pointed out that the anatomical characteristics of ACL in knee joints are different between

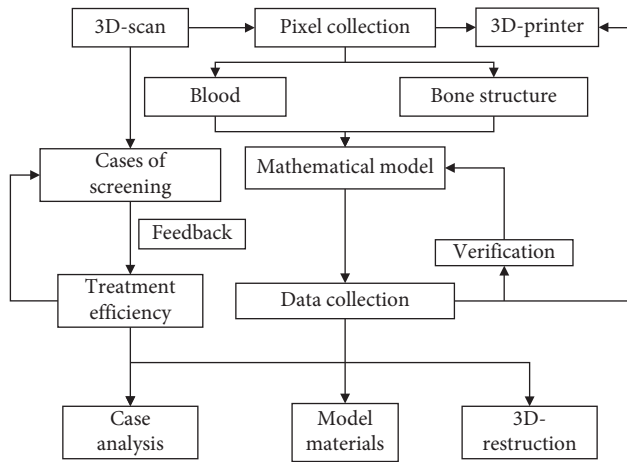


FIGURE 4: The schematic diagram for the process of 3D- reconstruction for bone tissues.

different races, genders, and ages. In this study, analysis of the ACL 3D reconstruction model data of subjects of different genders revealed that male ACL femoral stops were in terms of area, long-axis angle, distance, back distance, short axis, long axis, and other data. Both were significantly larger than females ( $P < 0.05$ ), suggesting that the impact of the patient's gender on the anatomy of the knee joint should be fully considered when formulating the knee joint reconstruction surgery under arthroscopy, so as to maximize the recovery of knee function of patients. In summary, there are anatomical characteristics of the knee joint anterior cruciate ligament femoral dead point, a certain individual difference. The combination of MRI and CI imaging data to achieve three-dimensional reconstruction can provide a reliable basis for the formulation of a surgical plan.

#### 4. Diagnosis and Exercise Rehabilitation of ACL Injuries

**4.1. Postoperative Recovery from Anterior Cruciate Ligament Injury.** Postoperative exercise rehabilitation is equally important in the treatment plan, and the specific process can be seen in Figure 6. Early scientific and reasonable exercise rehabilitation exercise treatment can promote graft healing and plastic reconstruction. At present, there is no consensus on the exercise rehabilitation treatment after anterior cruciate ligament reconstruction. It contains mainly conservative exercise rehabilitation treatment, intense exercise rehabilitation treatment, and individualized exercise rehabilitation treatment. Yasuda et al. advocate that the cycle of exercise rehabilitation is 1 year, and the patient starts the training of knee weight-bearing and activity in the second week after completing the anterior cruciate ligament reconstruction surgery. After training according to a conservative exercise rehabilitation treatment plan, the knee stability after reconstruction is good, but patients are prone to sequelae such as knee stiffness and poor joint mobility. Shellbourne et al. advocate that the treatment period for exercise rehabilitation is six months to nine months. Early in the postoperative period of anterior cruciate ligament

reconstruction, patients will begin training on knee weight bearing and mobility, which is a representative intense exercise rehabilitation treatment program. After exercise rehabilitation training, the patient can perform normal knee joint movements early after the operation; recovery of the knee joint activity is greater, and the incidence of knee complications can be reduced. However, some problems are prone to occur, such as loosened reconstructed grafts, even fractured grafts, and altered tunnels. The reason is that the degree of postoperative exercise rehabilitation activities is too strong, which may increase the relative activity of the graft in the bone canal. The Sharpey fibers formed early between the graft and the tunnel will be damaged, delaying the healing of the graft and the reconstruction stop, and even cause changes in the tunnel. The results of studies by foreign scholars show that patients with anterior cruciate ligament rupture reconstructed with autologous tendons (semiteudinosus and gracilis muscle) were divided into several groups of mutual control; one used a more intense exercise rehabilitation exercise treatment program, and for the other, a more conservative exercise rehabilitation exercise program was used for postoperative exercise rehabilitation, which is summarized in Table 2. On the plain radiograph of the knee joint 6 months after the operation, the changes in the tibial bone tract of the fierce exercise rehabilitation group were more obvious. The patients in the conservative exercise rehabilitation group had smaller changes in the tibial tunnel than the fierce exercise rehabilitation group. The intense exercise rehabilitation treatment will make the graft more active in the tunnel and, at the same time, will produce more severe cellular responses and inflammatory factors. With the development of sports medicine and exercise rehabilitation medicine, physicians have gradually realized that specific and effective exercise rehabilitation treatment programs must integrate all aspects of each patient to develop exercise rehabilitation training methods. In recent years, individualized exercise rehabilitation treatment programs have gradually started to rise, which requires professionals to be able to more accurately understand and master sports medicine and exercise rehabilitation medicine related knowledge. The patient first received a plain CT scan and then an MRI scan. The layer thickness was 2 mm, the layer spacing was 0.5 mm, the matrix size was set to  $512 \times 512$  pixels, and the number of excitations was taken twice. During the scan, the patient's knee joint was naturally straightened, and the external rotation was 10 to 15°. In order to enable the knee function to recover better and improve the shortcomings of conservative and intense exercise rehabilitation programs, exercise rehabilitation physicians need to formulate more individualized and targeted exercise rehabilitation treatment programs, which comprehensively analyze and consider all aspects of postoperative patients.

**4.2. The Role of Braces in Anterior Cruciate Ligament Injury.** In previous studies, knee functional braces were considered to be an integral part of exercise rehabilitation therapy, which can support the healing of ACL, improve the

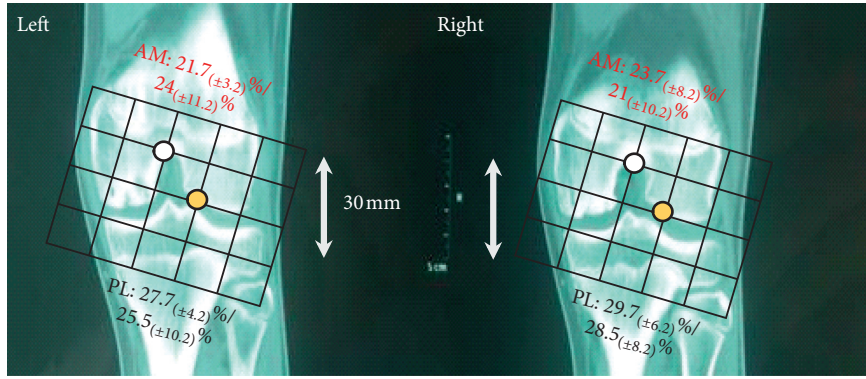


FIGURE 5: The ideal tibial bone tunnel positions for the left and right ACL femoral stops.

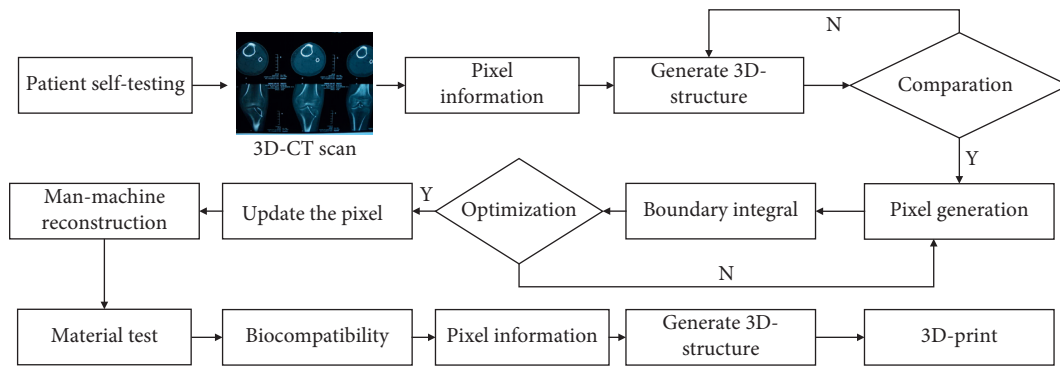


FIGURE 6: Analytical framework of application of 3.0 T MRI to knee joint sports injury assessment and exercise rehabilitation.

TABLE 2: Assessing grade results of knee joint sports injuries of patients in the 3.0 T MRI group ( $n$  (%)).

| Knee sports injury        | Level I<br>( $n$ %) | Level II<br>( $n$ %) | Level III<br>( $n$ %) |
|---------------------------|---------------------|----------------------|-----------------------|
| Meniscus injury           | 3 (30.00)           | 5 (50.00)            | 2 (20.00)             |
| Ligament injury           | 4 (33.33)           | 6 (50.00)            | 2 (16.67)             |
| Bone contusion            | 1 (12.50)           | 4 (50.00)            | 3 (37.50)             |
| Articular cavity effusion | 0 (0.00)            | 2 (66.67)            | 1 (33.33)             |

functional stability of joints, and restore normal kinematics. In different exercise rehabilitation stages, when performing exercise rehabilitation function exercises, the adjustable knee joint functional brace can adjust the knee joint activities in different ranges and has a supporting and protecting effect on the affected knee joint. In clinical exercise rehabilitation training after anterior cruciate ligament reconstruction, adjustable knee joint functional braces have been widely used. Here, the research angle could influence the growth rate of the bone section, so it could be easy to understand that it could also influence the kinematic function of the knee joint. However, there are still different opinions on the effects of postoperative exercise rehabilitation training on whether or not a brace should be worn. A systematic review summarizes the related research of 28 braces. The research angle is to consider whether the braces can improve the kinematic function of the knee joint. The

conclusion is that the literature does not currently recommend the use of knee function exercise rehabilitation after anterior cruciate ligament reconstruction braces, and it is believed that the use of knee functional exercise rehabilitation braces after anterior cruciate ligament reconstruction will not reduce postoperative pain and improve knee function. However, for patients after anterior cruciate ligament reconstruction, the main role of the knee functional exercise rehabilitation brace is first to support graft healing and maintain knee stability and then to perform functional exercise rehabilitation training. In addition, some scholars believe that if the postoperative patients do not wear braces for protection, knee function cannot be recovered well during exercise rehabilitation. It is recommended that patients who undergo anterior cruciate ligament surgery need to wear braces within 1 to 3 months after surgery. And the specific recovery comparison among different groups can be seen in Figure 7, and the details are summarized in Table 3. In the early stage of exercise rehabilitation, the brace should be mainly used to protect the knee joint. In the later stage of exercise rehabilitation, if there is a patient who needs to return to exercise, the choice of the brace should be mainly based on functional exercise.

Wearing a functional exercise rehabilitation brace after surgery can maintain knee stability, reduce postoperative pain, and protect the graft from injury without affecting the recovery of the patient's lower extremity strength. Although there is literature found that after the patient wears a brace

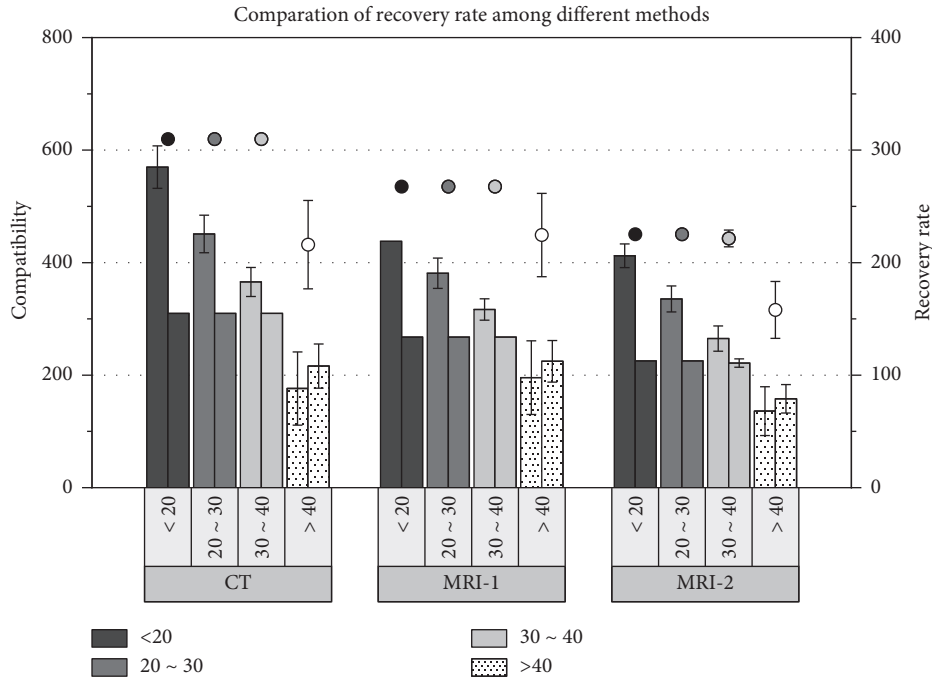


FIGURE 7: Compatibility and recovery rate at different recovery stages for different methods.

TABLE 3: Comparison of assessment results of different imaging methods on the exercise rehabilitation therapies of the three groups of patients.

| Group    | Level | <20        | 20~30      | 30~40      | >40        |
|----------|-------|------------|------------|------------|------------|
| CT       | 0     | 0 (0.00)   | 5 (12.50)  | 9 (22.50)  | 20 (50.00) |
|          | I     | 10 (25.00) | 13 (32.50) | 15 (37.50) | 12 (30.00) |
|          | II    | 15 (37.50) | 14 (35.00) | 12 (30.00) | 6 (15.00)  |
|          | III   | 13 (32.50) | 8 (20.00)  | 4 (10.00)  | 2 (5.00)   |
| MRI      | 0     | 0 (0.00)   | 3 (7.50)   | 8 (20.00)  | 14 (35.00) |
|          | I     | 11 (27.50) | 17 (42.50) | 16 (40.00) | 13 (32.50) |
|          | II    | 14 (35.00) | 11 (27.50) | 10 (25.00) | 9 (22.50)  |
|          | III   | 11 (27.50) | 9 (22.50)  | 6 (15.00)  | 4 (10.00)  |
| $\chi^2$ |       | 8.26       | 13.18      | 17.25      | 7.69       |
| $P$      |       | <0.05      | <0.05      | <0.05      | <0.05      |

after surgery, in terms of exercise and stress, the affected side has an inconsistency with the unoperated healthy side. In addition, there are reports in the literature that patients wearing braces may affect the sense of postoperative pain. This may increase the difficulty of gait training or motor skills for patients in the late stage of exercise rehabilitation. Moreover, there are no cases in the literature that report that wearing protective functional braces early after surgery will lead to changes in the reconstruction of the patient's tunnel. There are still different opinions on the use of braces, but no literature has found any reports of anterior cruciate ligament surgery. Patients wearing the brace will aggravate pain, and the recovery of knee joint mobility and graft healing and plastic reconstruction of ligaments after surgery are not affected. Although there are many types of functional exercise rehabilitation braces in the market and the quality and price are quite different, we have found that the general knee functional braces have knee motion trajectories that are

simply hinge flexion and extension. The normal knee joint is a series of complex three-dimensional movements accompanied by rolling and sliding of femoral condyles and axial internal rotation of tibia. Therefore, these knee joint braces currently used in China do not conform to the physiological structure of our human knee joints. According to reports in the literature, after anterior cruciate ligament injury, the tibia advances to the femur, and the severity of the injury varies, and the relative displacement of the tibia varies. The anterior cruciate ligament injury of the knee joint resulted in the tibia advancing to the femur.

## 5. Discussion

*5.1. Comparison of 3D-CT and MRI Mapping in Assessing ACL Cases.* There are many research methods about the ACL-related injuries of ACL femoral stops, mainly divided into several groups by different scan methods (Table 4). Through the comparison of the data in Figure 8, it is found that different methods can obtain different distortion reconstruction for bone tissues, even the same method, and the results obtained are different. The author believes that there are two reasons: (1) the positioning of Blumensaat lines by different authors is not very accurate; (2) there may be differences between different groups, populations, and individuals, leading to different positions of ACL femoral centers in this control group. For the selection of subjects, patients with high incidence of ACL injury were selected.

Compared with previous studies, it has the following advantages: (1) this study can do a large sample study, excluding the shortcomings of the limited number of studies on corpses, older age; (2) there may be individual differences in ACL anatomic location. Different races and populations may have different anatomical positions of ACL. This study

TABLE 4: Comparison of assessment results of different imaging methods on the exercise rehabilitation therapies of the three groups of patients.

| Reconstruction methods    | CT group |    |       | 1.5 T MRI group |    |       | 3.0 T MRI group |    |       |
|---------------------------|----------|----|-------|-----------------|----|-------|-----------------|----|-------|
|                           | 3.0 T    | AS | CR    | 1.5 T           | AS | CR    | CT              | AS | CR    |
| Meniscus injury           | 10       | 11 | 90.91 | 6               | 9  | 66.67 | 7               | 16 | 43.75 |
| Ligament injury           | 12       | 14 | 85.71 | 11              | 16 | 68.75 | 16              | 10 | 40.00 |
| Bone contusion            | 8        | 10 | 80.00 | 7               | 12 | 58.33 | 13              | 8  | 37.50 |
| Articular cavity effusion | 3        | 5  | 60.00 | 1               | 3  | 33.33 | 3               | 6  | 50.00 |

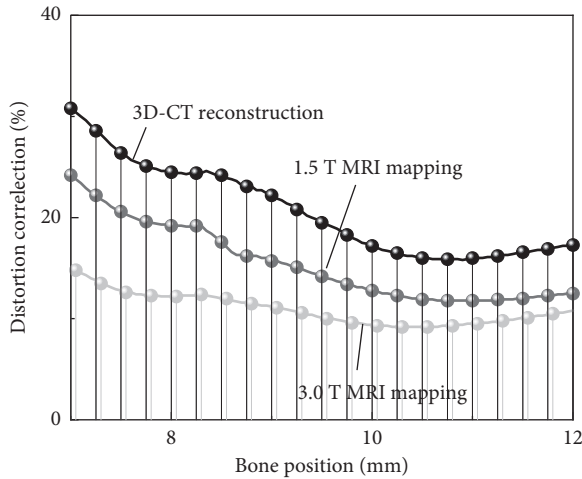


FIGURE 8: Distortion level for different reconstruction methods in at different recovery stages.

selected patients in the age group with high incidence of ACL injury in China, making the measured relative position of the center point more representative. The high-resolution DSC thin-layer scan used by the author can clearly show the natural imprint of the ACL femoral dead point and the position of the double-bone tract after surgery, making it possible to reproduce the anatomical position of the ACL dead point before surgery. Intraoperative and surgical positioning and postoperative evaluation of the position of the bone canal help to improve the surgical operation, and truly achieve the purpose of restoring the anatomy as much as possible. The volume reconstruction technology used can reconstruct the points and surfaces in any direction that the author is interested in, so that the imprint and bone can be visualized in front of the eyes, and overcome the lateral femoral condyle of interest that the author could not observe on the conventional X-ray plain film. However, the relative position of the bony structure was measured on the three-dimensional model, and the soft tissue of the ACL attachment point at the femoral end was not reconstructed. Due to the irregular imprint of the ACL stop, the group was drawn on the imprint circle and the center point. It is inevitable that there are individual subjective differences, as shown in Figure 9. In this study, the most commonly used quadrant method is used to compare the relative positions of the center points of the experimental group and the control group. The advantage is that the results of this study can be compared with the results of related reports in the past. However, the Blumensaat line is not visible on the 3D model

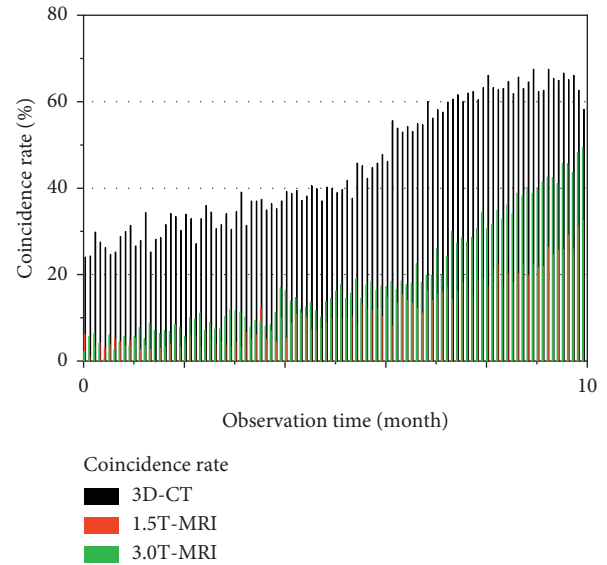


FIGURE 9: Coincidence rate for different reconstruction methods at different recovery stages.

of the image, and it is not a definite bone mark, as shown in Figure 10. Therefore, it is difficult to determine the Blumensaat line on the three-dimensional model. Different people inevitably have subjective individual differences when determining the position of the Blumensaat line.

**5.2. Biocompatibility of 3D-CT Modeling to Evaluate ACL and Related Cases.** The 3D-CT is a minimally invasive diagnostic method with high soft tissue resolution, which can clearly display bone lesions, and the characteristics of the images obtained by different sequences are different and can better identify and diagnose the type of injury. Using 3D-CT to diagnose knee joint sports injuries can obtain accurate diagnostic results and accurate classification, providing an accurate basis for forensic identification, as shown in Figure 11. Because the anatomical structure of the knee joint is more complex, the injury is wider, and the location of the injury is different, the impact on the function of the knee joint is also different. Therefore, in the process of identifying knee injuries, clinical forensics should analyze the patient's trauma history, injury mechanism, clinical characteristics, and imaging characteristics. In order to improve the accuracy of diagnosis, some injuries may cause complications because the knee joint is a weight-bearing joint and affected by factors such as improper function recovery training.



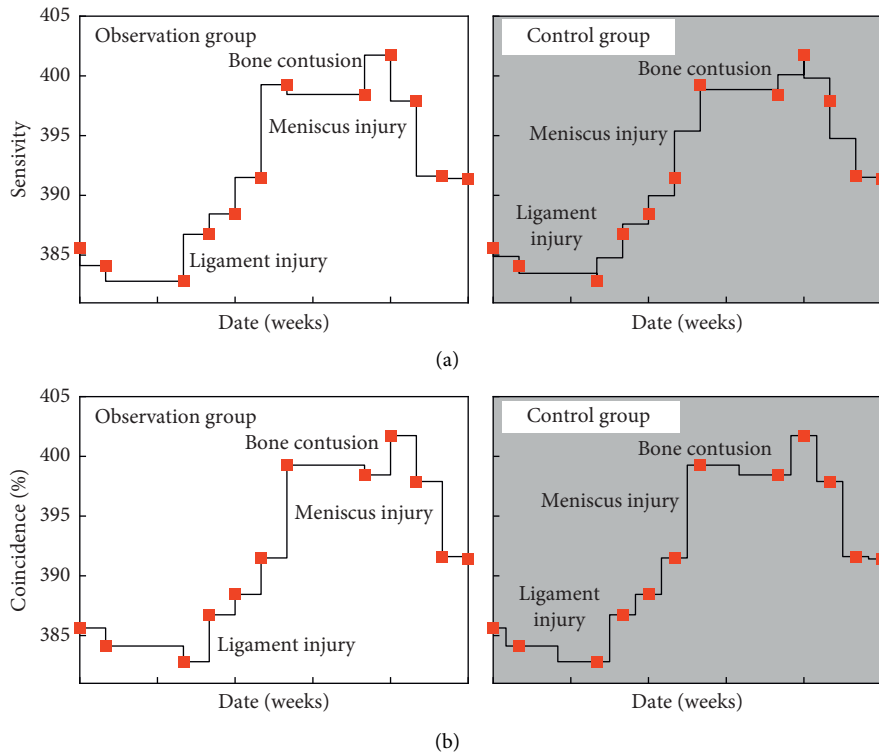


FIGURE 10: Sensitivity for different reconstruction methods at different recovery stages.

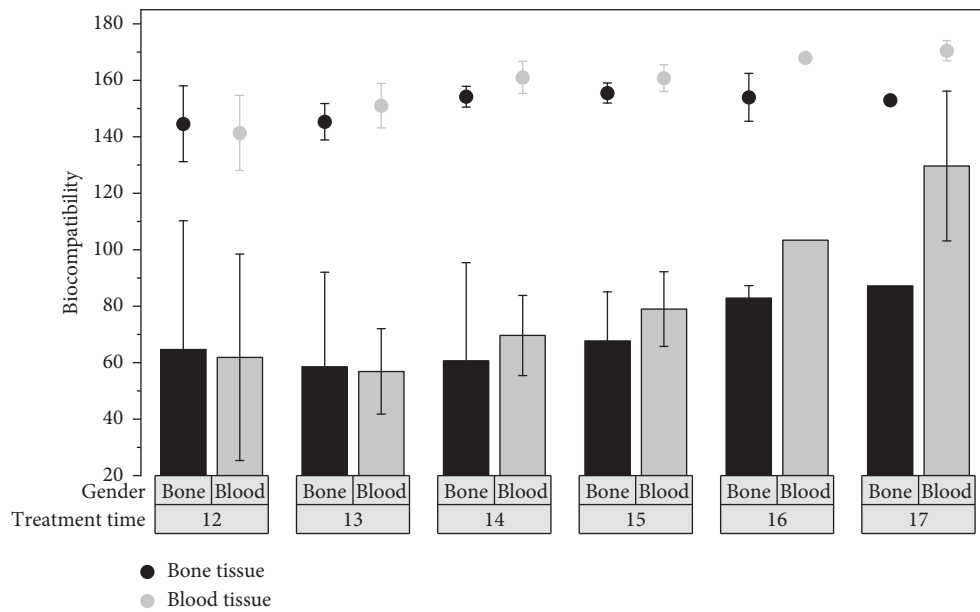


FIGURE 11: Biocompatibility of 3D-CT modeling to evaluate ACL and related cases in bone and blood tissues.

Therefore, when identifying them, they should be performed again after the patient's condition is stable and it is necessary to accurately grasp the best time for assessment, thereby ensuring the accuracy, objectivity, and impartiality of the conclusion of the assessment. In this study, all patients with anterior cruciate ligament injury were required to have the affected knee joint in an upright position and then take

X-rays from the side. Among them, the degree of advancement of the tibia relative to the femur was significantly different among different patients, and the tibia of some patients was significantly advanced relative to the femur. Later, it was reported in the literature that by scanning magnetic resonance imaging (MRI) of the knee joint, it was found that the degree of advancement of the lateral tibial

plateau to the lateral femoral condyle of patients with anterior cruciate ligament reconstruction failure was significantly increased. The results of this study further suggest that the clinical efficacy of anterior cruciate ligament reconstruction may be affected by the degree of advancement of the tibia relative to the femur after anterior cruciate ligament injury. Recently, a document reported by foreign scholars reported the role of soft tissues in maintaining the relationship between the normal position of the tibia and the femur, although the document confirmed the degree of forward displacement of the lateral tibia platform of the knee relative to the lateral femoral condyle in patients. When the anterior cruciate ligament is injured and the posterior horn of the lateral knee joint is damaged, it will increase significantly. However, there are no reports about the effect of bony structure on maintaining the relative position of tibia and femur after anterior cruciate ligament injury. At present, related literature has reported that one of the risk factors leading to anterior cruciate ligament injury is an increase in PTSA. Some literature studies have reported that through biomechanical studies, when the anterior cruciate ligament of the knee joint is injured, the displacement of the tibia relative to the femur will increase by about 6 mm forward with each increase of the tibial inclination of the tibia. Through the biomechanical model constructed by it, different foreign scholars have obtained similar experimental conclusions. However, at present, no scholars have proposed whether the relative position of the tibia relative to the femur will have different degrees of forward displacement after the anterior cruciate ligament injury in different patients.

## 6. Conclusion

In this paper, the 3D-CT three-dimensional reconstruction method is used to reconstruct the ACL natural femoral imprint and double-bone tract. The relative positional relationship between the two center points is compared, and the law is summarized to guide the improvement of ACL anatomic double-beam reconstruction under arthroscopy. The 3D reconstruction results suggest that the UF layer in the anterior medial portion is the thickest, forming a peak, and the thickness of the UF layer in the posterior medial portion gradually decreases in a stepwise manner. The entire UF tissue in the anterior medial portion and posterior medial portion is integrated into one body. The tissues are connected as a whole, and the thickness is relatively uniform. The two parts of the UF tissues are not connected. The CF tissue was inserted into the bone tissue in a zigzag pattern. The changes of CF tissues in the anterior medial and posteromedial CF tissues were similar, and they were distributed stepwise from the inside to the outside. According to the UF and CF spatial structure and changing rules, ACL is divided into medial and lateral beams. According to this study, it can be summarized that (1) 3D reconstruction can clearly reconstruct the natural footprint of ACL femoral stops and postoperative osseous position and (2) 3D reconstruction can be used to evaluate the position of osseous postoperative ACL anatomic double-beam reconstruction. In summary, the debate on how to reconstruct ACL will

continue, and the focus of everyone's debate is mainly on bone positioning and postoperative efficacy. Therefore, ACL surgery requires continuous improvement of surgical methods and surgical concepts.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The author declares that there are no conflicts of interest.

## References

- [1] G. Moatshe, J. Chahla, R. F. LaPrade, and L. Engebretsen, "Diagnosis and treatment of multiligament knee injury: state of the art," *Journal of ISAKOS: Joint Disorders & Orthopaedic Sports Medicine*, vol. 2, no. 3, pp. 152–161, 2017.
- [2] J. Chahla, I. R. Murray, J. Robinson et al., "Posterolateral corner of the knee: an expert consensus statement on diagnosis, classification, treatment, and rehabilitation," *Knee Surgery, Sports Traumatology, Arthroscopy*, vol. 27, no. 8, pp. 2520–2529, 2019.
- [3] E. H. Zadeh, "Diagnosis of meniscus tear of the knee by knee vibration signals processing," *Journal of Modern Rehabilitation*, vol. 9, no. 7, pp. 45–57, 2016.
- [4] M. F. Sobrado, "State of the art in multiligament knee injuries: from diagnosis to treatment," *Annals of Translational Medicine*, vol. 7, p. 23, 2019.
- [5] B. R. Deshpande, "Association of MRI findings and expert diagnosis of symptomatic meniscal tear among middle-aged and older adults with knee pain," *BMC Musculoskeletal Disorders*, vol. 17, no. 1, p. 154, 2016.
- [6] E. Mercuri, "Diagnosis and management of spinal muscular atrophy: part 1: recommendations for diagnosis, rehabilitation, orthopedic and nutritional care," *Neuromuscular Disorders*, vol. 28, no. 2, pp. 103–115, 2017.
- [7] D. J. Birnkrant, K. Bushby, C. M. Bann et al., "Diagnosis and management of duchenne muscular dystrophy, part 1: diagnosis, and neuromuscular, rehabilitation, endocrine, and gastrointestinal and nutritional management," *The Lancet Neurology*, vol. 17, no. 3, pp. 251–267, 2018.
- [8] A. A. Shmonin, V. M. Kasatkina, M. N. Maltseva, E. V. Melnikova, and G. E. Ivanova, "Analysis of problems in the rehabilitation diagnosis in the categories of the international classification of functioning in patients with stroke taking part in the pilot project "development of the system of medical rehabilitation in Russia,"" *Regional Blood Circulation and Microcirculation*, vol. 16, no. 2, pp. 17–24, 2017.
- [9] Z. Boychuck, A. Bussi eres, J. Goldschleger, and A. Majnemer, "Age at referral for diagnosis and rehabilitation services for cerebral palsy: a scoping review," *Developmental Medicine & Child Neurology*, vol. 61, no. 8, pp. 908–914, 2019.
- [10] S. Kohl, S. Meier, S. S. Ahmad et al., "Accuracy of cartilage-specific 3-tesla 3D-DESS magnetic resonance imaging in the diagnosis of chondral lesions: comparison with knee arthroscopy," *Journal of Orthopaedic Surgery and Research*, vol. 10, no. 1, p. 191, 2015.
- [11] W. Liu, H. Li, and Y. Hua, "Quantitative magnetic resonance imaging (MRI) analysis of anterior talofibular ligament in lateral chronic ankle instability ankles pre-and

- postoperatively," *BMC Musculoskeletal Disorders*, vol. 18, no. 1, p. 397, 2017.
- [12] Z. Li, J. Sun, X. Hu et al., "Assessment of liver fibrosis by variable flip angle T1 mapping at 3.0T," *Journal of Magnetic Resonance Imaging*, vol. 43, no. 3, pp. 698–703, 2016.
- [13] T. M. Noguerol, R. Barousse, M. Socolovsky, and A. Luna, "Quantitative magnetic resonance (MR) neurography for evaluation of peripheral nerves and plexus injuries," *Quantitative Imaging in Medicine and Surgery*, vol. 7, no. 4, p. 398, 2017.
- [14] A. C. Smith, K. A. Weber II, D. R. O'Dell, T. B. Parrish, M. Wasielewski, and J. M. Elliott, "Lateral corticospinal tract damage correlates with motor output in incomplete spinal cord injury," *Archives of Physical Medicine and Rehabilitation*, vol. 99, no. 4, pp. 660–666, 2018.
- [15] F. A. Chagas-Neto, M. H. Nogueira-Barbosa, M. M. Lorenzato et al., "Diagnostic performance of 3D TSE MRI versus 2D TSE MRI of the knee at 1.5 T, with prompt arthroscopic correlation, in the detection of meniscal and cruciate ligament tears," *Radiologia Brasileira*, vol. 49, no. 2, pp. 69–74, 2016.
- [16] B. E. Patterson, A. G. Culvenor, C. J. Barton et al., "Worsening knee osteoarthritis features on magnetic resonance imaging 1 to 5 years after anterior cruciate ligament reconstruction," *The American Journal of Sports Medicine*, vol. 46, no. 12, pp. 2873–2883, 2018.
- [17] A. M. Crespo-Rodríguez, J. C. De Lucas-Villarrubia, M. Pastrana-Ledesma, A. Hualde-Juvera, S. Méndez-Alonso, and M. Padron, "The diagnostic performance of non-contrast 3-tesla magnetic resonance imaging (3-T MRI) versus 1.5-tesla magnetic resonance arthrography (1.5-T MRA) in femoro-acetabular impingement," *European Journal of Radiology*, vol. 88, pp. 109–116, 2017.
- [18] J.-M. Kim, J.-M. Kim, B.-S. Jeon et al., "Comparison of postoperative magnetic resonance imaging and second-look arthroscopy for evaluating meniscal allograft transplantation," *Arthroscopy: The Journal of Arthroscopic & Related Surgery*, vol. 31, no. 5, pp. 859–866, 2015.
- [19] K. Okazaki, Y. Takayama, K. Osaki et al., "Subclinical cartilage degeneration in young athletes with posterior cruciate ligament injuries detected with T1 $\rho$  magnetic resonance imaging mapping," *Knee Surgery, Sports Traumatology, Arthroscopy*, vol. 23, no. 10, pp. 3094–3100, 2015.
- [20] A. J. Krych, D. H. Nawabi, N. A. Farshad-Amacker et al., "Bone marrow concentrate improves early cartilage phase maturation of a scaffold plug in the knee," *The American Journal of Sports Medicine*, vol. 44, no. 1, pp. 91–98, 2016.
- [21] K. Beattie, M. J. Davison, M. Noseworthy, J. D. Adachi, and M. R. Maly, "Quantifying fat and lean muscle in the lower legs of women with knee osteoarthritis using two different MRI systems," *Rheumatology International*, vol. 36, no. 6, pp. 855–862, 2016.
- [22] J. Zhang, L. Chen, Q.-X. Wang et al., "Isotropic three-dimensional fast spin-echo cube magnetic resonance dacryocystography: comparison with the three-dimensional fast-recovery fast spin-echo technique," *Neuroradiology*, vol. 57, no. 4, pp. 357–365, 2015.
- [23] S. Liu, Y. Sun, T. Chen et al., "Time from injury to surgery affects graft maturation following posterior cruciate ligament reconstruction with remnant preservation: a magnetic resonance imaging-based study," *Arthroscopy: The Journal of Arthroscopic & Related Surgery*, vol. 34, no. 10, pp. 2846–2854, 2018.
- [24] R. Y. Huang, L. Tian, B. G. Xu, X. Z. Ma, and Y. Zhang, "Magnetic resonance imaging analysis of knee joint injury of college football players in Xianyang city," *Clinical Medical Research and Practice*, vol. 6, pp. 26–130, 2019.
- [25] X. N. Zheng, X. Fu, and Y. X. Wu, "Application of 3.0 T MRI multiple sequence imaging in knee joint trauma," *Henan Journal of Surgery*, vol. 20, no. 4, pp. 25–27, 2014.
- [26] D. E. Hartigan, K. W. Carroll, F. J. Kosarek, D. P. Piasecki, J. F. Fleischli, and D. F. D'Alessandro, "Visibility of anterolateral ligament tears in anterior cruciate ligament-deficient knees with standard 1.5-tesla magnetic resonance imaging," *Arthroscopy: The Journal of Arthroscopic & Related Surgery*, vol. 32, no. 10, pp. 2061–2065, 2016.
- [27] A. Matsushita, T. Maeda, E. Mori et al., "Can the acute magnetic resonance imaging features reflect neurologic prognosis in patients with cervical spinal cord injury?" *The Spine Journal*, vol. 17, no. 9, pp. 1319–1324, 2017.
- [28] Y. Qiu, J. Yao, X. Wu et al., "Longitudinal assessment of oxytocin efficacy on bone and bone marrow fat masses in a rabbit osteoporosis model through 3.0-T magnetic resonance spectroscopy and micro-CT," *Osteoporosis International*, vol. 26, no. 3, pp. 1081–1092, 2015.
- [29] A. Santos, M. Vicente, J. de Brito, I. Flores-Colen, and A. Castelo, "Inspection, diagnosis, and rehabilitation system of door and window frames," *Journal of Performance of Constructed Facilities*, vol. 31, no. 3, Article ID 04016118, 2017.
- [30] R. D. Zorowitz, "ICU-acquired weakness: a rehabilitation perspective of diagnosis, treatment, and functional management," *Chest*, vol. 150, no. 4, pp. 966–971, 2016.
- [31] J. Conceição, B. Poça, J. de Brito, I. Flores-Colen, and A. Castelo, "Data analysis of inspection, diagnosis, and rehabilitation of flat roofs," *Journal of Performance of Constructed Facilities*, vol. 33, no. 1, Article ID 04018100, 2019.
- [32] L. Hubermann, Z. Boychuck, M. Shevell, and A. Majnemer, "Age at referral of children for initial diagnosis of cerebral palsy and rehabilitation: current practices," *Journal of Child Neurology*, vol. 31, no. 3, pp. 364–369, 2016.
- [33] Y. Gu, J. S. Baker, I. M. Lanzoni, and Y.-T. Lim, "A special section on medical imaging techniques for sports injuries diagnosis and rehabilitation," *Journal of Medical Imaging and Health Informatics*, vol. 10, no. 5, pp. 1133–1135, 2020.

## Research Article

# Big Data Analytics for Complex Credit Risk Assessment of Network Lending Based on SMOTE Algorithm

Aiwen Niu,<sup>1</sup> Bingqing Cai,<sup>2</sup> and Shousong Cai <sup>3</sup>

<sup>1</sup>Glorious Sun School of Business Management, Donghua University, Shanghai 200051, China

<sup>2</sup>School of Humanities, Shanghai University of Finance and Economics, Shanghai 200433, China

<sup>3</sup>School of Business Administration, Shanghai Lixin University of Accounting and Finance, Shanghai 201209, China

Correspondence should be addressed to Shousong Cai; [caishousong@126.com](mailto:caishousong@126.com)

Received 8 July 2020; Accepted 4 August 2020; Published 26 September 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Aiwen Niu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the continuous development of big data technology, the data of online lending platform witness explosive development. How to give full play to the advantages of data, establish a credit risk assessment model, and realize the effective control of platform credit risk have become the focus of online lending platform. In view of the fact that the network loan data are mainly unbalanced data, the smote algorithm is helpful to optimize the model and improve the evaluation performance of the model. Relevant research shows that stochastic forest model has higher applicability in credit risk assessment, and cart, ANN, C4.5, and other algorithms are also widely used. In the influencing factors of credit evaluation, the weight of the applicant's enterprise scale, working years, historical records, credit score, and other indicators is relatively high, while the index weight of marriage and housing/car production (loan) is relatively low.

## 1. Introduction

In recent years, with the rapid development of Internet finance, P2P platform develops rapidly in this context and gradually forms a new financial platform with great influence. P2P platform relies on cloud computing, social networking, and other channels to collect, organize, and record data, which greatly enhances the ability of financial risk prevention and control based on data mining technology. Through comparative analysis of user information, combining with specific historical data, it can effectively improve the information flow efficiency of both sides of capital supply and demand and provide necessary support for both sides to establish the relationship between supply and demand, and based on this, the financial risk caused by information asymmetry is kept at the lowest level [1]. However, the domestic P2P platform research time is relatively short, has not yet established a sound credit system, coupled with the relevant legal system is not perfect, and is easy to induce credit risk, which is a serious threat to the safety of users'

funds. In addition, with the advent of the era of big data, the data of online lending platform are constantly increasing, the data types are diverse, and the data are updated quickly [2]. How to give full play to the data advantages, obtain the required information, and enhance the platform's ability to monitor capital risk factors become the key to the development of the platform. In this case, the platform needs to rely on big data, combined with data mining technology, to build a scientific and reasonable credit risk assessment model to provide necessary basis for platform risk supervision and user investment. Therefore, this study is of great significance on the practical level.

Relatively speaking, the development time of foreign P2P platform is relatively long, and the level of related research results is high, which has more reference value in the aspects of concept discussion and risk assessment [3]. In the aspect of loan success rate, the relationship between personal information and loan success rate has been studied, and then the borrowing strategies are comprehensively discussed with the help of quantitative analysis tools. In terms of credit risk,

domestic scholars use empirical analysis and specific cases to analyze the influencing factors of credit risk and summarize the influencing factors related to default behavior [4]. Based on this, a classification method with random forest as the core is constructed, which greatly improves the effectiveness of credit risk assessment [5]. Affected by the Internet, computer, information technology, etc., building a smart city has become a key task of socialist construction work [6]. The empirical analysis shows that compared with FICO or LC, the evaluation method based on random forest has more advantages in identifying high reputation borrowers [7]. The research results in recent years show that the role of social network cannot be ignored in the development of online lending. The higher the richness of social resources, the lower the cost of obtaining loans [8]. There is a significant negative correlation between the two [9]. Empirical analysis shows that through the analysis of applicants' social networks, we can deeply understand the soft information related to credit risk so as to evaluate the applicants' credit risk more comprehensively. In the dual-channel supply chain system, channel optimization is influenced by channel attitude toward risk, in which risk is classified as general risk and interruption risk [10]. For individuals, P2P platform provides convenience for its financing or capital problems but also produces a series of risk problems, such as imperfect credit system, high moral hazard, and serious adverse selection [11]. At present, credit risk is the key content of risk research, and the research direction includes default characteristics analysis and platform reputation [12]. As the number of selected lines increases, the current same price for all passengers in different riding paths could make the bus industry development a step further [13]. In terms of default probability of applicants, incomplete market-oriented interest rate has more significant prediction effect, but the use of personal public information can also reflect the default risk to a certain extent [14]. Taking Renrendai as an example, its credit certification mechanism has certain advantages in reflecting credit risk, but the index system has certain limitations, so it is necessary to supplement and improve its evaluation index system.

Overall, the depth of research on P2P network lending needs to be further expanded. Compared with western developed countries, China's P2P platform development time is relatively short and mainly concentrated after the rise of Internet finance in 2012. The empirical research data are insufficient, mainly referring to the data provided by foreign platforms. However, both the research methods and the research conclusions are difficult to fully meet the domestic research demand. In view of this situation, this paper introduces R language and python to write web crawler program in data crawling of online credit platform, introduces smote algorithm in unbalanced data processing, and constructs credit risk assessment model combined with six data mining algorithms, which is more consistent with the development of domestic P2P platform. It is a kind of research on network credit risk based on big data background, and the new ideas play a positive role in improving the level of theoretical research in China.

## 2. Model Description

### 2.1. Commonly Used Data Mining Classification Model.

In the field of classification technology, decision tree presents the classification process in the form of directed acyclic tree, which is intuitive and simple, so it has a high application rate [15]. For classified data, the greedy algorithm is used as the core of decision tree to determine the nodes, and then the local optimal decision strategy is used to construct the decision tree. In the dual-channel supply chain system, channel optimization is influenced by channel attitude toward risk, in which risk is classified as general risk and interruption risk [16]. There are significant differences in decision tree types with different classification criteria. For example, taking information theory as the standard, it can be divided into ID3, C4.5, and cart, and SLIQ and sprint can be obtained from Gini index. Among the above methods, only ID3 can be used for discrete variables. On the basis of comprehensive analysis, cart and C4.5 algorithms are selected in this paper.

AdaBoost algorithm is a kind of lifting algorithm which can adjust the distribution of training samples by itself. It has high adaptive ability to ensure that the base classifier fits the samples with higher difficulty in classification. Through the AdaBoost algorithm, the weights of training samples can be combined, the parameters can be updated, and then the corresponding weighting can be completed:

$$\omega_i^{(j+1)} = \frac{\omega_i^{(j)}}{Z_j} \times \begin{cases} e^{-\alpha_j}, C_j(X_i) = y_i, \\ e^{\alpha_j}, C_j(X_i) \neq y_i, \end{cases} \quad (1)$$

where  $\omega_i^{(j)}$  refers to the weight of the sample  $(X_i, y_i)$  in the round of  $j$  iteration. Using this weight can enhance the weight of the wrong classification samples to a certain extent, which is not conducive to highlighting the weight of the correct classification samples [17]. Therefore, for unbalanced data sets, this algorithm can improve the accuracy of minority prediction to the greatest extent, and its defect is that the fitting problem is more prominent.

Support vector machine (SVM) is a realization method based on statistical learning theory. This method relies on the Mercer theorem and combines with nonlinear mapping method to realize the effective mapping of feature space in the Hilbert space and realize the accurate division of samples according to the linear decision boundary [18]. The application fields of this method include nonlinear regression model, high-level data analysis, and sample classification.

Artificial neural network (ANN) is a method to analyze the law of things by imitating the organizational structure of biological neural network. It is based on a large number of nodes with connection relationship, which can realize continuous iteration by connecting different nodes. The online-to-offline (O2O) business model is the new online shopping model in which consumers purchase products or services online and get the products or services in offline physical stores [19]. In this process, we need to determine the weight of the previous iteration, then calculate the weight of the node, and update the weight with the error value. Through the repetition of the above process, the error is reduced to the allowable range. Practice

shows that neural network is suitable for sample classification and variable regression and has good application effect. However, due to the high sensitivity of this method to noise, it is prone to local minimum problem, which has a certain negative impact on the accuracy of the final results.

**2.2. Random Forests.** Random forest is a combined classifier algorithm with decision tree as its core. In this method, the cart algorithm is used to construct the decision tree, the decision tree is used as the metaclassifier of sample classification, and the corresponding training set is obtained. In the construction of a single decision tree, the corresponding variables can be determined randomly, and node splitting can be completed based on the vector. According to the characteristics of this method, random forest has high robustness to noise, but low sensitivity to multiple reproducibility, so it can be relatively robust to deal with nonequilibrium data and get reasonable results.

The core of random forest is tree classifier, which is composed of various types of classifiers  $\{h(x, \theta_k), k = 1, 2, \dots, n\}$ . Among them, the nonconstructed classification decision tree obtained by using the cart algorithm is the metaclassifier. According to the simple arithmetic average of single decision tree and majority voting output results, the accurate result data can be achieved, and the steps are as follows.

Firstly, the training sample set is constructed. In general, self-help resampling technology can be used to generate independent sample sets; that is, based on  $n$  sample sets,  $k$  new organizational sample sets are obtained by random return, and then the corresponding decision tree is formed, while the unselected samples constitute out-of-bag data, namely, OOB.

Secondly, the decision tree node is split. According to the overall situation of decision tree characteristic variables, assume  $m$  and then randomly determine  $m$  characteristic variables from them to split the corresponding nodes. Among them, the number of characteristic variables randomly obtained by each node is less than the number of assumed characteristic variables, and the corresponding splitting is carried out according to the principle of node impure minimization. It should be emphasized that all decision trees have no pruning operation.

Thirdly, the decision tree completes the corresponding combination. Based on the decision trees obtained in the above steps, the output results are determined by averaging all decision trees by majority voting, and then the error analysis stage is entered.

For the data of nontraining set, the possibility of error classification by a specific classifier is the generalization error. Theoretical research shows that if the number of decision trees reaches a certain degree, the upper bound of random forest generalization error will converge according to the law of large numbers. Under the premise of the given sample, the interval function provided by using the random forest is as follows:

$$mr(x, y) = P_{\Theta}(h(x, \theta_k) = y) - \max_{j \neq y} P_{\Theta}(h(x, \theta_k) = j). \quad (2)$$

The strength of classifier set  $\{h(x, \theta)\}$  can be expressed as follows:

$$s = E_{X,Y}mr(x, y). \quad (3)$$

According to the above expression, there is a positive correlation between the strength of the classifier set and the value of the interval function, that is, the strength of the classifier set increases with the increase of the value of the interval function, and the prediction accuracy will also be improved accordingly:

$$PE^* \leq \frac{\bar{\rho}(1 - s^2)}{s^2}. \quad (4)$$

According to the above expression, the upper bound of generalization error is negatively correlated with the strength of the combined classifier, but positively correlated with the decision tree. Therefore, by weakening the correlation or enhancing the strength of a single decision tree, the generalization error performance can be improved.

The first is the OOB estimation. The bagging method can be used in self-service sampling. If the data are selected in the future, it will be used to predict the classification accuracy, that is, the OOB estimation of classification error rate. After averaging, the random forest generalization error estimation can be obtained.

The second is the characteristic importance value. The application of the random forest method can determine the specific degree of the importance of a single eigenvalue. At the same time, the performance of each decision tree can be evaluated by using the data outside the bag, that is, the accuracy rate of OOB. By combining with the noise interference test, the performance of the decision tree can be tested more accurately, that is, the new OOB accuracy rate. The important value of feature  $V$  in the decision tree can be expressed as the difference between the accuracy of new and old OOB, and its important value is determined after averaging. If there are more features in the basic samples, the best model can be determined by sorting the important values. Figure 1 shows the parameter selection of the random forest algorithm.

### 3. Data Collection and Processing

**3.1. Data Sources.** According to the relevant data, there are more than 1700 domestic P2P platforms, which complete the lending process with the help of third-party platforms. At present, Renrendai is the largest and longest established P2P platform in China. Therefore, this paper selects Renrendai loan as the research object, combines R language and python to write a web crawler program, obtains its relevant data, and gets about 50 variables, including amount and interest rate.

#### 3.2. Data Preprocessing

*Step 1.* Eliminate the variables that do not meet the conditions. Specifically, it includes the variables with the same values, the variables with repeated specific contents, the variables not related to the research topic, and the variables with serious missing data.

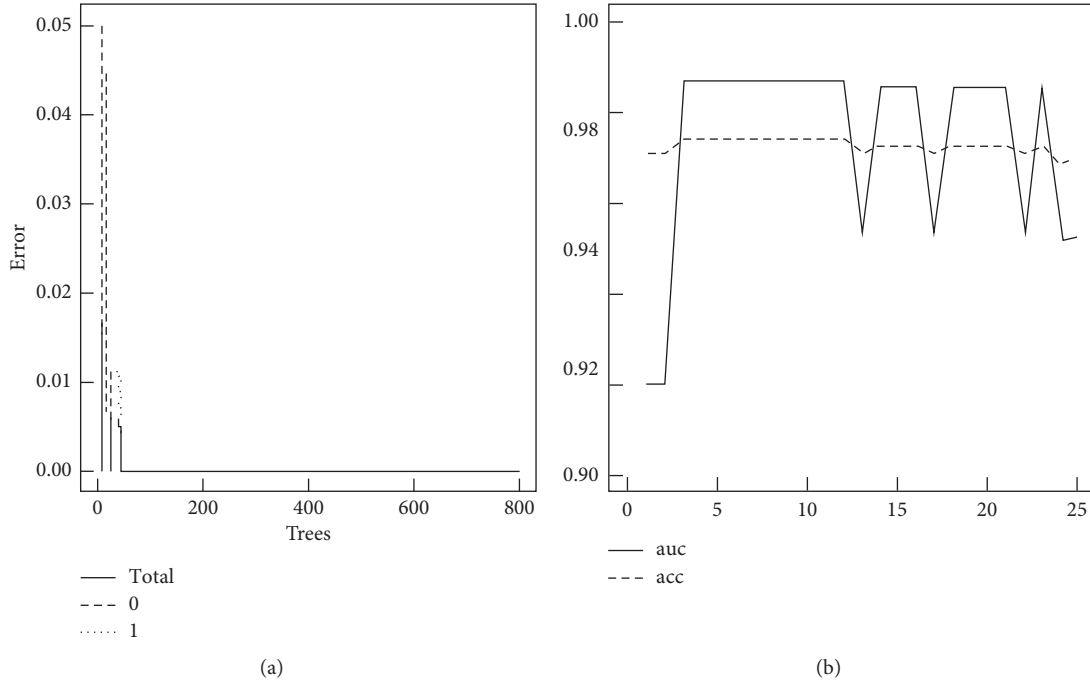


FIGURE 1: Parameter selection of the random forest algorithm. (a) ntree. (b) mtry.

*Step 2. Missing value processing.* It is found that the variables of some loan items are incomplete, such as the lack of industry, enterprise scale, and position. According to the specific situation, its industry can be defined as e-commerce; the enterprise scale is expressed as 0; and the position is expressed as individual shopkeeper.

*Step 3. Data normalization processing.* The output variable selects the number of overdue times, in which if it exceeds 0, it is marked as 0; otherwise, it is marked as 1; 0 and 1 are used to represent binary variable values; integers are used to represent education level, subject type, etc.; working hours are represented by the median value; and the amount of loan can be expressed as  $x' = (X - \min(X)) / (\max(X) - \min(X)) \times 10$ . The basic information of data after preprocessing is shown in Table 1.

#### 4. Credit Risk Assessment Model Based on Data Mining Algorithm

*4.1. Unbalanced Data Processing.* In the data sample obtained in this paper, there are 30 default items, accounting for 2.935%, and the rest are nondefault items, namely, unbalanced data set. Traditional data mining algorithms have some limitations in dealing with unbalanced distribution classes, and it is difficult to effectively focus on a few classes. Therefore, its classification performance is difficult to meet the requirements.

The data sampling method can be selected, that is, up or down sampling; on the contrary, data mining algorithms can be optimized and improved, such as cost sensitive learning. Through comparative analysis, it can be found that the possibility of incomplete data in down sampling is high.

Therefore, the application of upward sampling is more extensive. The basic up sampling method achieves the balance of data sets by randomly copying a few samples, but it is difficult to avoid the fitting problem.

The smote algorithm uses a small number of samples to construct artificial samples, thus achieving the balance of data sets, which is conducive to avoid the over fitting phenomenon. In this algorithm, the artificial samples are inserted into the adjacent samples in the feature space to increase the number of samples. For  $X_i \in S_{\min}$ ,  $k$  nearest neighbor points are searched, and the nearest neighbor points are selected by combining the parameters such as correlation coefficient or Euclidean distance. After determining the nearest neighbor points, the corresponding sample points  $Y_j$  are determined. On the basis of determining the difference between  $X_i$  and  $Y_j$  with the corresponding eigenvector, the random number  $\delta$  is determined, and then the artificial samples  $X_{\text{new}}$  are determined as follows:

$$X_{\text{new}} = X_i + (X_i - Y_j) \times \delta, \quad (5)$$

where  $j$  is the number of sample points  $j = 1, 2, \dots, n$ . Repeat the above steps, and stop after all minority samples are processed.

The smote algorithm uses the smote function to complete the confirmation of majority class sample  $m$  and minority class  $n$ . On the basis of determining the up and down sampling rates, the final majority class sample and minority class sample number  $N + nN$  and  $nNm$  are obtained.

The first is sample classification, which is divided into the test set sample and training set sample by random sampling; the second training set data balance is that minority class

TABLE 1: Basic information of data.

| No. | Variables   | Actual meaning            | Data processing   |
|-----|-------------|---------------------------|---|
| 1   | ID          | User ID                   |   |
| 2   | BorrowType  | Type of standards         | 1: credit certification standard; 2: institutional guarantee standard; 3: field certification standard  |
| 3   | Title       | Loan title                | 1: capital turnover; 2: expansion of production/operation; 3: decoration; 4: purchase of goods/raw materials/equipment; 5: daily life consumption; 6: purchase of cars; 7: investment in entrepreneurship; 8: purchase of house; 9: online store stocking/business expansion/capital turnover; 10: education and training expenditures; 11: marriage expenses; 12: others |
| 4   | Amount      | Loan amount               | 0~10  |
| 5   | Interest    | Interest rate             |   |
| 6   | Term        | Repayment period          |   |
| 7   | LoanType    | Way of guarantee          | 1: principal; 2: principal + interest   |
| 8   | Prepayment  | Prepayment                |   |
| 9   | Gender      | Gender                    | 0: female; 1: male  |
| 10  | Age         | Age                       |   |
| 11  | Education   | Education                 | 1: high school or below; 2: junior college; 3: undergraduate; 4: graduate or above  |
| 12  | Marriage    | Marriage                  | 1: married; 2: unmarried; 3: divorced; 4: widowed   |
| 13  | Field       | Company industry          | 0: one; 1: retail/wholesale; 2: manufacturing; 3: public utilities; 4: catering/hotel industry; 5: construction engineering; 6: nonprofit organization; 7: education/training; 8: transportation industry; 9: IT; 10: finance/law; 11: medical/hygiene/health; 12: government agencies; 13: others  |
| 14  | Size        | Company scale             | 0: one; 1: <10 persons; 2: 10–100 persons; 3: 100–500 persons; 4: >500 persons  |
| 15  | EmpLength   | Years of working          | 1: 1 year or less; 2: 1–3 years (including 3 years); 3: 3–5 years (including 5 years); 4: more than 5 years   |
| 16  | Income      | Income                    | 12000~5000 RMB; 25000~10000 RMB; 310000~20000 RMB; 420000~50000 RMB; more than 550000 RMB   |
| 17  | House       | House                     | 0: no; 1: yes   |
| 18  | Mortgage    | Mortgage                  | 0: no; 1: yes   |
| 19  | Car         | Car                       | 0: no; 1: yes   |
| 20  | Autoloan    | Car loan                  | 0: no; 1: yes   |
| 21  | Score       | Credit score              |   |
| 22  | Grade       | Credit rating             | 1:A; 2: B; 3: C; 4: D; 5: E; 6: IIR   |
| 23  | Application | Borrowing application     |   |
| 24  | Succeed     | Successful borrowing      |   |
| 25  | Paid        | Number of paying off      |   |
| 26  | Total       | Total amount of borrowing | 0~10  |
| 27  | Overdue     | Overdue times             |   |
| 28  | Status      | Repayment situation       | 0: overdue time >0; 1: overdue time = 0   |

$N = 15$ , majority class  $M = 496$ , taking  $n = 500\%$ ,  $m = 200\%$ ,  $k = 5$ , and keeping the proportion of 3 : 5, so as to improve the model performance. Table 2 shows the data composition.

**4.2. Model Empirical Analysis.** In this paper, the classification variable is repayment, and then the scientific selection of model parameters is performed, in order to obtain the analysis results of different data mining models, to lay the foundation for the subsequent empirical analysis. Table 3 lists the results of parameter selection and important variables of each model.

- (1) Determine the model parameters and output the corresponding results. In this paper, the random forest algorithm is selected to determine the number of decision trees and the number of variables of node branches (mtry). Then the model is built according to the new training set. If the number of decision trees is less than 40, the error rate fluctuation is not more

than 0.05; if the number of decision trees is more than 40, the prediction error rate is reduced to 0; determine the selected variables of 3–13 nodes to achieve the maximum AUC and accuracy under stable state. To sum up, ntree = 800 and mtry = 3 can be selected to complete the model construction, and each category can be accurately predicted. Figure 2 shows the Friedman average ranking.

On the whole, the variables with higher importance were paid, succeeded, application, score, field, etc., while the variables with lower importance were house and marriage. The importance of some variables is 0. Therefore, in the process of credit risk assessment, personal work information, credit rating, and historical records are the main variables. Relatively speaking, the importance of personal life information is lower than the above variables. Taking Renrendai as an example, the platform is based on the credit rating mechanism, combined with the



TABLE 2: Data composition.

| Data processing | Data set                  | 0  | 1   | Total | 0 (%) | 1 (%)  |
|-----------------|---------------------------|----|-----|-------|-------|--------|
|                 | Original data             | 30 | 992 | 1022  | 2.935 | 97.065 |
| Random sampling | Test set (test)           | 15 | 496 | 511   | 2.935 | 97.065 |
|                 | Training set (train)      | 15 | 496 | 511   | 2.935 | 97.065 |
| SMOTE           | New training set (ntrain) | 90 | 150 | 240   | 37.5  | 62.5   |

TABLE 3: Parameter selection and important variables of each model.

| Model    | Original training set  |   | New training set  |   |
|----------|--|---|---|---|
|          | Parameter selection  | Important variables                             | Parameter selection   | Important variables                               |
| CART     | —  | Succeed, field                                  | —   | Succeed   |
| C4.5     | —  | Succeed, field, size                            | —   | Autoloan, score                                   |
| AdaBoost | —  | Succeed, field, tile, application, size, score  | —   | Succeed, empLength, paid, size, grade, borrowType |
| SVM      | C: classification mode, polynomial kernel function                                 | —   | C: classification mode, polynomial kernel function, weight is 2:1.4               | —   |
| ANN      | Number of hidden nodes = 6, maximum iteration times = 200                          | —   | Number of hidden nodes = 11, maximum iteration times = 207                        | —   |
| RF       | Number of spanning trees = 800, number of variables selected by node branches = 25 | Paid, succeed, score, field, application, grade | Number of spanning trees = 800, number of variables selected by node branches = 3 | Paid, succeed, application, score, size, grade    |

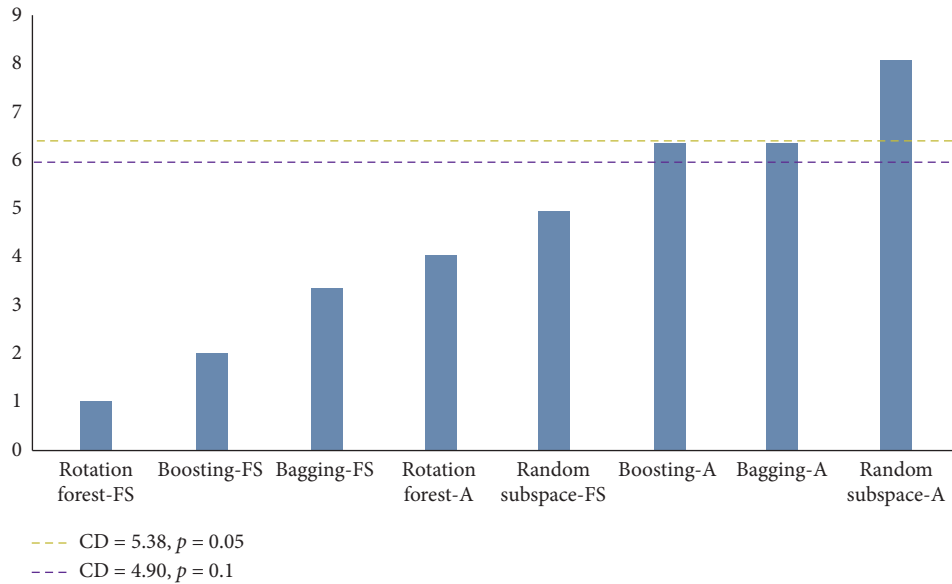


FIGURE 2: Friedman average ranking.

materials provided by the applicant, and serves as a reference for investors. In the main variables of credit risk evaluation, historical loan information can reflect the use of customer loans, while personal work information can reflect the stability of applicants' work, which is an important reference for evaluating their repayment ability. For the platform, we must further strengthen the collection, collation, and storage of data, provide more powerful information support for credit risk assessment and qualification review, and help investors obtain more income on

the premise of ensuring the safety of investors' funds to the greatest extent. Table 4 shows classification results summary of each model.

- (2) Comparative analysis of model performance is done before and after data balance processing. In the aspect of classifier performance evaluation, the accuracy index is usually selected; however, for the classification of unbalanced data, it is not appropriate to select the accuracy only. Therefore, the original model can be optimized by combining the

TABLE 4: Classification results summary of each model.

| Model    | Original training set |          |                    |                    |               |       | New training set |        |                    |                    |               |       |       |       |
|----------|-----------------------|----------|--------------------|--------------------|---------------|-------|------------------|--------|--------------------|--------------------|---------------|-------|-------|-------|
|          | 0                     | 1        | True positive rate | True negative rate | Accuracy rate | AUC   | 0                | 1      | True positive rate | True negative rate | Accuracy rate | AUC   |       |       |
| CART     | 0<br>1                | 13<br>12 | 2<br>481           | 0.7                | 0.977         | 0.972 | 0.885            | 0<br>1 | 13<br>17           | 1<br>476           | 0.932         | 0.962 | 0.962 | 0.948 |
| C4.5     | 0<br>1                | 8<br>13  | 5<br>483           | 0.5                | 0.975         | 0.964 | 0.784            | 0<br>1 | 12<br>10           | 1<br>481           | 0.866         | 0.971 | 0.972 | 0.921 |
| AdaBoost | 0<br>1                | 8<br>7   | 5<br>488           | 0.5                | 0.985         | 0.972 | 0.792            | 0<br>1 | 11<br>13           | 3<br>480           | 0.7           | 0.971 | 0.963 | 0.882 |
| SVM      | 0<br>1                | 7<br>2   | 6<br>492           | 0.531              | 0.992         | 0.97  | 0.761            | 0<br>1 | 12<br>11           | 4<br>479           | 0.732         | 0.975 | 0.961 | 0.854 |
| ANN      | 0<br>1                | 9<br>8   | 4<br>485           | 0.666              | 0.981         | 0.972 | 0.825            | 0<br>1 | 11<br>4            | 3<br>489           | 0.7           | 0.983 | 0.981 | 0.891 |
| RF       | 0                     | 9        | 4                  | 0.665              | 0.997         | 0.987 | 0.833            | 0      | 14                 | 0                  | 2             | 0.971 | 0.972 | 0.984 |

TABLE 5: Results of 3-fold cross validation.

| Model    | 3-fold cross validation | TP | FP | FN | TN  | True positive rate | True negative rate | Accuracy rate | AUC  |
|----------|-------------------------|----|----|----|-----|--------------------|--------------------|---------------|------|
| CART     | 1                       | 9  | 10 | 0  | 321 | 9                  | 0966               | 0965          | 0982 |
|          | 2                       | 6  | 4  | 2  | 325 | 08                 | 0984               | 0972          | 0841 |
|          | 3                       | 8  | 14 | 0  | 314 | 08                 | 0952               | 0954          | 0928 |
|          | Mean value              |    |    |    |     | 0866               | 0965               | 0961          | 0917 |
| C4.5     | 1                       | 8  | 8  | 2  | 321 | 08                 | 0972               | 0972          | 0935 |
|          | 2                       | 6  | 2  | 2  | 327 | 06                 | 0994               | 0983          | 0841 |
|          | 3                       | 8  | 11 | 0  | 317 | 08                 | 0962               | 0963          | 0933 |
|          | Mean value              |    |    |    |     | 0832               | 0975               | 0973          | 0906 |
| AdaBoost | 1                       | 8  | 7  | 2  | 321 | 08                 | 0975               | 0975          | 0939 |
|          | 2                       | 6  | 5  | 4  | 324 | 06                 | 0981               | 0975          | 0842 |
|          | 3                       | 6  | 8  | 4  | 320 | 06                 | 0972               | 0965          | 0835 |
|          | Mean value              |    |    |    |     | 0762               | 0975               | 0972          | 0871 |
| SVM      | 1                       | 6  | 11 | 4  | 318 | 06                 | 0963               | 0952          | 0835 |
|          | 2                       | 6  | 10 | 4  | 321 | 06                 | 0966               | 0956          | 0835 |
|          | 3                       | 7  | 18 | 3  | 310 | 07                 | 0941               | 0934          | 0875 |
|          | Mean value              |    |    |    |     | 0732               | 0955               | 0952          | 0841 |
| ANN      | 1                       | 7  | 11 | 3  | 318 | 07                 | 0961               | 0953          | 0883 |
|          | 2                       | 8  | 14 | 2  | 317 | 08                 | 0954               | 0954          | 0928 |
|          | 3                       | 8  | 23 | 2  | 307 | 08                 | 0926               | 0925          | 0915 |
|          | Mean value              |    |    |    |     | 0866               | 0942               | 0942          | 0911 |
| RF       | 1                       | 9  | 6  | 1  | 325 | 9                  | 0971               | 0981          | 0983 |
|          | 2                       | 6  | 2  | 4  | 327 | 06                 | 0992               | 0981          | 0841 |
|          | 3                       | 9  | 11 | 1  | 317 | 9                  | 0963               | 0964          | 0983 |
|          | Mean value              |    |    |    |     | 0911               | 0977               | 0972          | 0935 |

specific index and sensitivity index. We can compare the ROC curve of each model in Figure 3.

The first is accuracy. The model built by the new training set can basically achieve the accuracy rate of 0.963–0.982, among which ANN, RF, and C4.5 rank in the top three. Even though the accuracy of cart, AdaBoost, SVM, RF has declined, it can more accurately predict a few items. Among them, C4.5 and ANN models have greatly improved the prediction accuracy based on the original training set model.

The second is ROC curve and AUC. The closer to the upper left corner, the higher the accuracy of the model prediction. Comparatively speaking, the ROC

curve of the model constructed based on the new training set is more concentrated in the upper left corner, which indicates that the classifier has better performance. In particular, after the original sample is balanced by using the smote algorithm, the model constructed based on this has significantly higher AUC, which is more than 0.85. RF, cart, and C4.5 rank in the top three, and RF, cart, and C4.5 rank the best. The AUC of random forest method is very close to 1, reaching 0.987. Compared with other models, its advantages are very significant.

Generally speaking, in the related research of credit risk evaluation, it is of great significance to

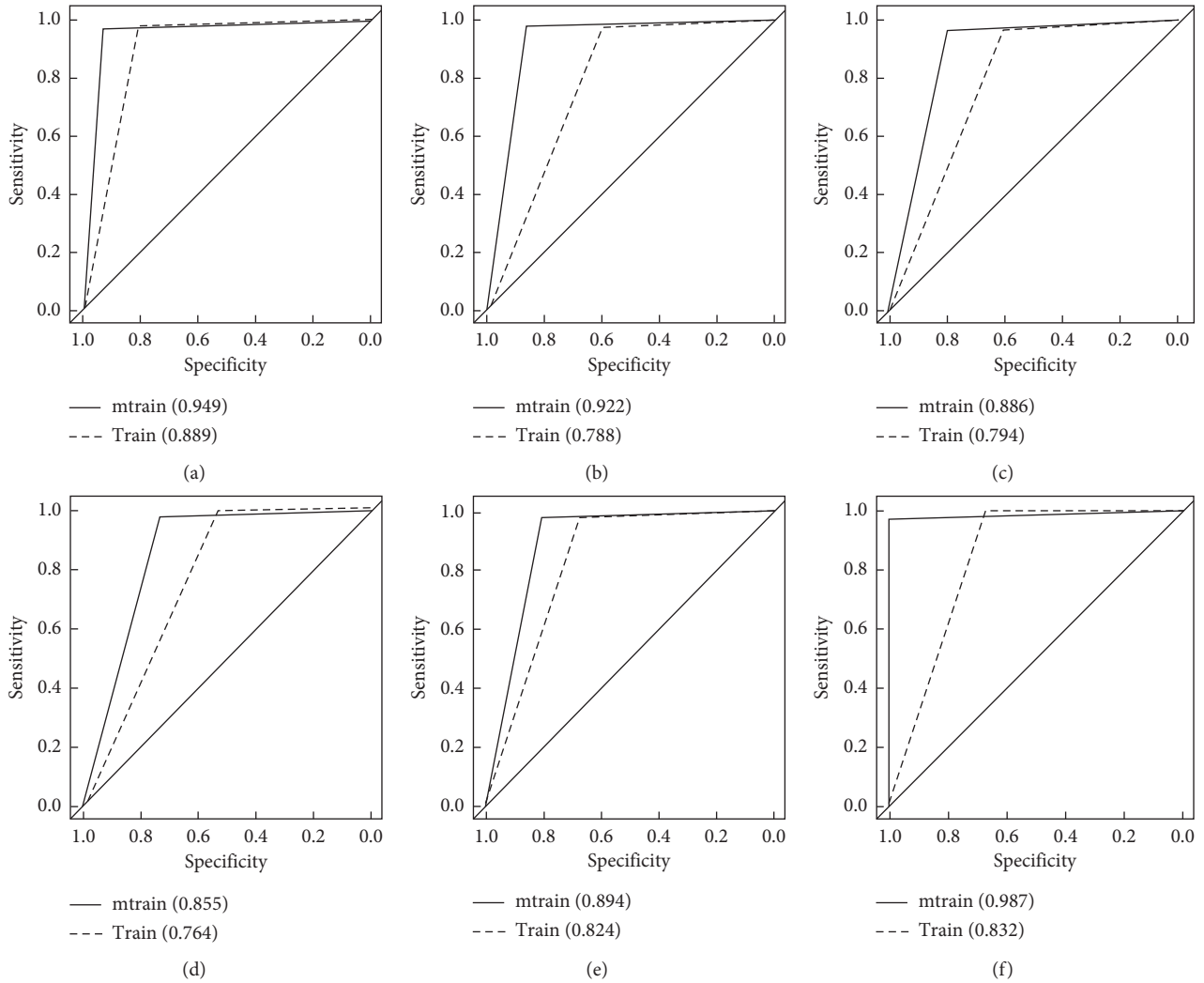


FIGURE 3: ROC curve of each model. (a) CART. (b) C45. (c) AdaBoost. (d) SVM. (e) RF. (f) ANN.

strengthen the research on the prediction of a small number of samples, which can provide more information support for relevant investors, help investors choose investment projects more scientifically, so as to minimize the credit risk, and improve the security of funds, which has good practical value. In this case, according to the characteristics of the original training set, this paper introduces the smote algorithm to deal with it, which greatly improves the performance of credit risk assessment model, and improves the accuracy of default project prediction.

- (3) The prediction performance of different models was compared and analyzed. Through the analysis of Table 4, it can be found that the true rate of the random forest model in the model built based on the new training set is about 1, and its AUC is as high as 0.987, which has relatively high accuracy and has good identification ability for relevant default samples. In summary, this paper preliminarily

determines that the random forest model has higher prediction accuracy and the best performance.

In order to verify the conclusion of this paper and determine the best model, this paper selects a 3-fold cross validation method. According to the standard of this paper, dependent variables include default variables and nondefault variables. In order to balance the two categories in the original data, we can divide them into three parts randomly, that is, three data sets including default variables and nondefault variables, and run them as test sets. The data sets are processed by using the smote algorithm, and then the corresponding models are established and their classification performance is evaluated to carry out the targeted test. It can be seen from Table 5 that the mean values of true positive rate are larger, the difference is larger.

Among them, the models with more than 0.85 include RF, cart, and ANN, which are in the forefront. Therefore, the above three models have high recognition ability for default items; the true negative rate of RF, AdaBoost, and C4.5 is in the top three, and the accuracy rate of RF, C4.5, and

AdaBoost is in the top three. Considering that the accuracy rate is difficult to distinguish the minority class from the majority class, the accuracy can only be used as a reference to determine the best model, rather than the main factor. The AUC of RF, cart, and ANN ranked in the top three. To sum up, the best performance is the random forest model, which has broad application prospects in the evaluation of network lending credit.

## 5. Conclusion

This paper comprehensively and systematically studies the credit risk factors in P2P network lending and constructs a data mining model in risk assessment, which lays the foundation for the follow-up research. The smote algorithm is used to process the unbalanced data, and then the corresponding model is established, which can reduce the volatility of prediction accuracy and improve the risk identification ability of AUC index and default items. The future research focuses on the following: first, strengthen the analysis of user behavior; second, judge the correlation between user behavior and credit risk; and third, build a user credit risk assessment system to provide real-time search function for the platform.

## Data Availability

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The findings were sponsored by the National Social Science Fund of China (Grant no. 18CGL015).

## References

- [1] H. Liang, J. Zou, K. Zou, and M. J. Khan, "An improved genetic algorithm optimization fuzzy controller applied to the wellhead back pressure control system," *Mechanical Systems and Signal Processing*, vol. 142, no. 8, Article ID 106708, 2020.
- [2] J. Hu, Y. Sun, G. Li, G. Jiang, and B. Tao, "Probability analysis for grasp planning facing the field of medical robotics," *Measurement*, vol. 141, no. 7, pp. 227–234, 2019.
- [3] F. Hu and G. Wu, "Distributed error correction of EKF algorithm in multi-sensor fusion localization model," *IEEE Access*, vol. 8, no. 5, pp. 93211–93218, 2020.
- [4] D. Jiang, G. Li, Y. Sun, J. Kong, and B. Tao, "Gesture recognition based on skeletonization algorithm and CNN with ASL database," *Multimedia Tools and Applications*, vol. 78, no. 21, pp. 29953–29970, 2019.
- [5] H. Liang, J. Zou, Z. Li, M. J. Khan, and Y. Lu, "Dynamic evaluation of drilling leakage risk based on fuzzy theory and PSO-SVR algorithm," *Future Generation Computer Systems*, vol. 95, no. 6, pp. 454–466, 2019.
- [6] Z. Liu, B. Hu, B. Huang, L. Lang, H. Guo, and Y. Zhao, "Decision optimization of low-carbon dual-channel supply chain of auto parts based on smart city architecture," *Complexity*, vol. 2020, no. 5, 14 pages, Article ID 2145951, 2020.
- [7] M. Buchholz and L. Tonzer, "Sovereign credit risk comovements in the eurozone: simple interdependence or contagion?" *International Finance*, vol. 19, no. 3, pp. 246–268, 2016.
- [8] W. Sun, Y. Zhao, and L. Sun, "Big data analytics for venture capital application: towards innovation performance improvement," *International Journal of Information Management*, vol. 50, no. 2, pp. 557–565, 2020.
- [9] M. Xie, H. Li, and Y. Zhao, "Blockchain financial investment based on deep learning network algorithm," *Journal of Computational and Applied Mathematics*, vol. 372, no. 7, p. 112723, 2020.
- [10] Z. Liu, B. Hu, Y. Zhao et al., "Research on intelligent decision of low carbon supply chain based on carbon tax constraints in human-driven edge computing," *IEEE Access*, vol. 8, no. 3, pp. 48264–48273, 2020.
- [11] L. Sun, Y. Zhao, and W. Sun, "Study on supply chain strategy based on cost income model and multi-access edge computing under the background of the internet of things," *Neural Computing and Applications*, 2019.
- [12] S. Morris and H. S. Shin, "Illiquidity component of credit risk—the 2015 Lawrence R. Klein lecture," *International Economic Review*, vol. 57, no. 4, pp. 1135–1148, 2016.
- [13] Z. Liu, S. Chen, B. Hu, M. Zhou, and Y. Zhao, "Research on staged pricing model and simulation of intelligent urban transportation," *IEEE Access*, vol. 7, no. 9, pp. 141404–141413, 2019.
- [14] F. Wang, L. Ding, H. Yu, and Y. Zhao, "big data analytics on enterprise credit risk evaluation of e-business platform," *Information Systems and e-Business Management*, 2019.
- [15] J. Wang, X. Wang, F. Meng, R. Yang, and Y. Zhao, "Massive information management system of digital library based on deep learning algorithm in the background of big data," *Behaviour & Information Technology*, pp. 1–12, 2020.
- [16] L. Zheng, X. Qi, and K. Yang, "Optimal independent pricing strategies of dual-channel supply chain based on risk-aversion attitudes," *Asia-Pacific Journal of Operational Research*, vol. 35, no. 2, pp. 1–17, 2018.
- [17] E. Ramentol, Y. Caballero, R. Bello, and F. Herrera, "SMOTE-RSB\*: a hybrid preprocessing approach based on over-sampling and undersampling for high imbalanced data-sets using SMOTE and rough sets theory," *Knowledge and Information Systems*, vol. 33, no. 2, pp. 245–265, 2012.
- [18] R. Yang, L. Yu, Y. Zhao et al., "Big data analytics for financial market volatility forecast based on support vector machine," *International Journal of Information Management*, vol. 50, no. 2, pp. 452–462, 2020.
- [19] X. Qi, W. Wang, and L. Zheng, "The influence of online subsidies service on online-to-offline supply chain," *Asia-Pacific Journal of Operational Research*, vol. 35, no. 2, pp. 1–14, Article ID 1840007, 2018.

## Research Article

# Articulatory-to-Acoustic Conversion Using BiLSTM-CNN Word-Attention-Based Method

Guofeng Ren , Guicheng Shao, and Jianmei Fu

Department of Electronics, Xinzhou Teachers University, Xinzhou 034000, China

Correspondence should be addressed to Guofeng Ren; [renguofeng926@sina.com](mailto:renguofeng926@sina.com)

Received 6 July 2020; Revised 27 August 2020; Accepted 12 September 2020; Published 26 September 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Guofeng Ren et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the recent years, along with the development of artificial intelligence (AI) and man-machine interaction technology, speech recognition and production have been asked to adapt to the rapid development of AI and man-machine technology, which need to improve recognition accuracy through adding novel features, fusing the feature, and improving recognition methods. Aiming at developing novel recognition feature and application to speech recognition, this paper presents a new method for articulatory-to-acoustic conversion. In the study, we have converted articulatory features (i.e., velocities of tongue and motion of lips) into acoustic features (i.e., the second formant and Mel-Cepstra). By considering the graphical representation of the articulators' motion, this study combined Bidirectional Long Short-Term Memory (BiLSTM) with convolution neural network (CNN) and adopted the idea of word attention in Mandarin to extract semantic features. In this paper, we used the electromagnetic articulography (EMA) database designed by Taiyuan University of Technology, which contains ten speakers' 299 disyllables and sentences of Mandarin, and extracted 8-dimensional articulatory features and 1-dimensional semantic feature relying on the word-attention layer; we then trained 200 samples and tested 99 samples for the articulatory-to-acoustic conversion. Finally, Root Mean Square Error (RMSE), Mean Mel-Cepstral Distortion (MMCD), and correlation coefficient have been used to evaluate the conversion effect and for comparison with Gaussian Mixture Model (GMM) and BiLSTM of recurrent neural network (BiLSTM-RNN). The results illustrated that the MMCD of Mel-Frequency Cepstrum Coefficient (MFCC) was 1.467 dB, and the RMSE of F2 was 22.10 Hz. The research results of this study can be used in the features fusion and speech recognition to improve the accuracy of recognition.

## 1. Introduction

Along with the popularity of artificial intelligence, man-machine interaction technology has put forward higher requirements for speech processing technology, and it is hoped that intelligent products, such as computers and mobile phones, will have the ability to communicate harmoniously with human beings and the ability to express emotions. The existing technology of emotional speech processing inevitably took advantage of the human pronunciation mechanism, and then human speech is pronounced successfully by the systematic movements through the muscle's contraction of the vocal organs, such as the tongue, lips, and jaw. This relationship between articulatory and acoustic data has been formed through the accumulation of a great deal of articulatory experience.

Although people have adopted a variety of technologies to collect the motion information of articulators, such as X-ray [1], real-time Magnetic Resonance Imaging (rMRI) [2], Ultrasound [3], EPG [4], and EMA [5], most data acquisition environments were not perfect, and the collected data were of poor natural degree or were easily disturbed by external noise [6]. Among them, due to the EMA technology using sensors placed on the pronunciation organs such as the surface of the lip, contact area is only 3 mm<sup>2</sup>; at the same time, the sensors' working theory is simple and with stable performance, which has been widely used in pronunciation organs' trajectory tracking and data collection.

For more than a decade, researchers have been studying the acoustic-to-articulatory inversion. Ouni and Laprie [7] first proposed the codebook method in 2005, which used vector quantization to encode the acoustic vectors of speech

and calculate the minimal Euclidean distance between the acoustic vectors and the articulatory vectors, so as to construct the inversion system. The drawback of this method is that it requires a large amount of data to achieve the accurate conversion effect.

King and Wrench [8] implemented a dynamic system to train EMA data using Kalman filter in 1999. They defined the acoustic and articulatory features of speech as linear relationship based on the physical model of speech production. However, there was no strict linear relationship between the acoustic and articulatory features.

Furthermore, in 2000, Dusan and Deng [9] used an extended Kalman filter to train the acoustic-articulatory data to establish a more realistic inversion relationship. By combining this model with Kalman smoothing filter, the movement trajectory of the articulator would be simulated, and the RMSE between the simulated trajectory and the original trajectory was realized to be 2 mm.

Korin Richmond and Yamagishi [10] used neural network to realize the acoustic-to-articulatory inversion firstly in 2002. They used the data of two subjects in MOCHA-TIMIT and achieved the inversion result with a RMSE as low as 1.40 mm. At the same time, Toda et al. [11] proposed a feature inversion method based on Gaussian Mixture Model (GMM), which used maximum likelihood estimation method to analyze the parallel acoustic data stream and the EMA data stream, and established the joint probability density function. Different quantities of Gaussian mixture elements had been used to achieve higher inversion accuracy.

Hiroya and Honda [12], Lin et al. [13], and Ling et al. [14] successively used and improved HMM and finally achieved the integrated RMS of 1.076 mm, which is also the highest inversion accuracy achieved by using HMM model so far.

In recent years, deep learning has attracted great attention for its ability to model nonlinear mapping relations and has been applied to the inversion of articulatory and acoustic features. Leonardo Badino et al. [15, 16] realized acoustic-to-articulatory inversion using Deep Belief Network (DBN) and Hidden Markov Model (HMM) and applied it to speech recognition, resulting in a 16.6% reduction in the recognition relative bit error rate. At the early stage, convolutional neural network (CNN) [17] has been widely used in the field of image signal processing, which had obvious advantages in the analysis of local features; meanwhile the articulatory features could be seen as the visual features of speech. Sun et al. [18] from Yunnan University showed that CNN could be applied to the emotion classification of speech and achieved good results. They were the first to introduce word-attention mechanism to emotion classification and reveal the influence of semantics on classification effect.

However, most researchers only focus on the acoustic-to-articulatory inversion, and the research on the articulatory-to-acoustic conversion is less and started relatively late. Yet the articulatory-to-acoustic conversion is helpful to the study of pronunciation mechanism and the development of speaker recognition and emotion recognition. Liu et al. [19, 20] of the University of Science and Technology of

China used Cascade Resonance Network and BiLSTM-RNN to convert articulatory features into spectral energy and fundamental frequency features in 2016 and 2018, respectively, and achieved a good conversion effect.

At present, the conversion focuses on the frame or phoneme level, with emphasis on the pronunciation rules and acoustic characteristics of phonemes. However, in the tonal languages like Mandarin, the interaction between syllables must hide certain acoustic-pronunciation information. Meanwhile, word-attention mechanism has been widely applied in the field of text processing and emotion classification. Wang and Chen [21] proposed an LSTM emotion classification method based on attention mechanism and realized emotion classification through feature screening of short- and long-text features combined with attention mechanism. Wang et al. [22] proposed a word-attention convolution model with the combination of CNN and attention mechanism, aiming at word feature extraction.

Relying on deep learning with nonlinear and attention mechanism, BiLSTM-CNN method and word-attention mechanisms were used to realize the articulatory-to-acoustic conversion in this paper. The paper is organized as follows. First, we review related work on articulatory-to-acoustic conversion, as well as CNN and word-attention mechanism in Section 2. Next, the detailed method we proposed is described in Section 3, and Section 4 reports our experiments and their results. Section 5 provides the discussion and conclusion of the work.

## 2. Related Work

To explore the articulatory-to-acoustic conversion and improve the conversion effect, lots of researches have been carried out in the past decades, and several methods have been proposed to model the conversion, including Gaussian Mixture Model (GMM), recurrent neural network (RNN), Long Short-Term Memory (LSTM), BiLSTM, and CNN. We will give a brief introduction in this section.

*2.1. GMM-Based Articulatory-to-Acoustic Conversion.* GMM is a classical feature conversion method [23], which used the joint probability density function of acoustic-articulatory features to realize the conversion. The description of the transformation model is

$$y_i = \sum_{i=1}^M p(\lambda_i | t_i) p(y_i | x_i, \lambda_i), \quad (1)$$

Here,  $M$  is used to represent the number of Gaussian mixture elements,  $p(\lambda_i | x_i)$  denotes the probability of acoustic feature vector  $x_i$ , and  $p(y_i | x_i, \lambda_i)$  represents full covariance matrix of conditional Gaussian distribution.

$x = (x_1, x_2, \dots, x_M)$  and  $y = (y_1, y_2, \dots, y_M)$  have been defined as articulatory and acoustic features, respectively, where,  $M$  is the number of frames. Considering that the articulatory features of frame  $i$  are known, the first-order dynamic features are as follows:

$$\Delta x_i = -0.5x_{i-1} + 0.5x_{i+1}. \quad (2)$$

The articulatory features and the first-order dynamic features are spliced as the input feature vector  $X_i = [x_i^M, \Delta x_i^M]^M$ , and then output vector  $Y_i = [y_i^M, \Delta y_i^M]^M$  can be obtained. Thus, the joint probability distribution of input and output vectors can be described as follows:

$$P(Z_i | \vartheta) = P(X_i, Y_i | \Theta) = \prod_{j=1}^N \partial_j \mathbb{N} \left( Z_i, \mu_j^Z, \sum_j^Z \right), \quad (3)$$

$$Z_i = \begin{bmatrix} X_i \\ Y_i \end{bmatrix},$$

$$\mu_j^Z = \begin{bmatrix} \mu_j^X \\ \mu_j^Y \end{bmatrix},$$

$$\sum_j^Z = \begin{bmatrix} \sum_j^{XX} & \sum_j^{XY} \\ \sum_j^{YX} & \sum_j^{YY} \end{bmatrix},$$

where  $Z_i = [X_i^M, Y_i^M]$  is the joint vector of articulatory and acoustic features,  $N$  is the number of Gaussian elements,  $\theta = \alpha_j, \mu_j^Z, \sum_j^Z | j = 1, 2, \dots, N$ , denotes the model parameters of GMM, and  $\alpha_j, \mu_j^Z$ , and  $\sum_j^Z$  are weight, mean, and covariance of Gaussian element  $j$ , respectively. Among them, model parameter  $\Theta$  will be estimated by Maximum Likelihood Estimation Algorithm (MLEA) [24]. When the dimension between articulatory and acoustic features is different, the covariance matrix  $\sum_j^Z$  is full-rank matrix.

During the conversion, input articulatory features are supposed to be  $X = [x_1, x_2, \dots, x_M]$ , and output acoustic features are supposed to be  $Y = [y_1, y_2, \dots, y_M]$ ;  $y^*$  can be calculated relying on the MLE as follows:

$$y^* = \arg \max_y p(Y | X, \Theta), \quad \text{s.t. } Y = W y, \quad (4)$$

Here,  $W$  is dynamic window coefficient matrix. In formula (4), conditional probability distribution can be rewritten as follows:

$$p(Y | X, \Theta) = \sum_{j=1}^N p(j | X, \Theta) p(Y | j, Y, \Theta). \quad (5)$$

If we only refer to a Gaussian element, it can be calculated by Maximum Posterior Probability, which is shown as follows:

$$j^* = \arg \max p(j | X, \Theta). \quad (6)$$

If the frames are independent of each other, for input of frame  $i$ ,  $X_i$  exist as Formula (7); meanwhile, the output of frame  $i$ ,  $Y_i$  exist as Formula (8):

$$p(j | X_i, \Theta) = \frac{p(j | \Theta) p(X_i | j, \Theta)}{p(X_i | \Theta)} = \frac{\alpha_j \mathbb{N}(X_i | \mu_j^X, \sum_j^{XX})}{\sum_n \alpha_n \mathbb{N}(X_i | \mu_n^X, \sum_n^{XX})}, \quad (7)$$

$$p(Y_i | j^*, X_i, \Theta) = \mathbb{N}(Y_i | \mu_{j^*}^{y|x}, \sum_{j^*}^{y|x}). \quad (8)$$

Here,  $\mu_{j^*}^{y|x}$  and  $\sum_{j^*}^{y|x}$  are mean and covariance matrix, respectively, which are calculated using the following two formulas:

$$\mu_{j^*}^{y|x} = \mu_{j^*}^Y + \sum_{j^*}^{YX} \sum_{j^*}^{YX^{-1}} (X_i - \mu_{j^*}^X), \quad (9)$$

$$\sum_{j^*}^{y|x} = \sum_{j^*}^{YY} - \sum_{j^*}^{YX} \sum_{j^*}^{XX} \sum_{j^*}^{XY}. \quad (10)$$

On this basis, we can obtain the output sequence using maximum likelihood criterion, as shown in formula (11), where  $\Sigma^y$  is square matrix and  $\mu^y$  can be calculated through  $\mu_{j^*,t}^{y|x}$  connecting nose to tail:

$$y^* = \left( W^T \sum_{j^*}^{y-1} W \right)^{-1} W^T \sum_{j^*}^{y-1} \mu^y. \quad (11)$$

**2.2. LSTM of RNN.** Recurrent neural network (RNN) is a kind of neural network that takes sequence data as input data and recurses along the time domain direction of the sequence [20]. All nodes in this network are connected in a chain. RNN has the advantages of memorability, parameter sharing, and Turing completeness and is obviously superior to GMM in learning nonlinear features. The network has been widely used in speech recognition, speech modeling, feature conversion, and other fields.

The core of RNN is the directed graph, and the loop unit is fully connected. Input sequence is given as  $X = \{X_1, X_2, \dots, X_M\}$ , and spread length is given as  $\tau$ . For time-step  $t$ , the recurrent unit should be taken as

$$h^{(t)} = f(s^{(t-1)}, X^{(t)}, \theta), \quad (12)$$

where  $h$  denotes systematic state of RNN,  $s$  denotes inner state calculated by  $s = s(h, X, y)$ , and  $f$  represents activation function, such as logistic and hyperbolic tangent function, or represents a kind of feedforward neural network. The excitation function corresponds to the simple recurrent network, and the feedforward neural network corresponds to some depth algorithms.  $\theta$  is weight coefficient in the recurrent unit.

We take the example of an RNN containing a hidden layer; the hidden layer vector sequence  $H = \{h_1, h_2, \dots, h_M\}$  can be obtained by

$$h^{(t)} = f(W_{hh} h^{(t-1)} + W_{xh} X^{(t)} + \theta_h). \quad (13)$$

Then, output sequence  $Y = \{Y_1, Y_2, \dots, Y_M\}$  can be shown as follows:

$$Y^{(t)} = W_{hy} h^{(t)} + \theta_y. \quad (14)$$

Initially, inverse error transfer algorithm on the time axis has been adopted to update the parameters, which would produce some inverse transfer error. So gradient erasing and explosion would occur, which seriously affected the training effect of RNN. In order to reduce the above problems, Li et al. [25] put forward Long Short-Term Memory (LSTM), including nonlinear transform and gate-structure affection function. Through the development of LSTM, the structure brought forward by Aviles and Kouki [26] is consisting of

input gate, output gate and forgetting gate. Among them, input gate is used to control the conversion processing from accepted information to memory sequence, which is shown as follows:

$$i^{(t)} = \sigma(W_{ix}X^{(t)} + W_{ih}h^{(t-1)} + W_{ic}c^{(t)} + \theta_i), \quad (15)$$

Here,  $\sigma$  is sigmoid function and  $c$  is memory sequence. Forgetting gate is used to control how much of the current memory information should be discarded, the implementation method of which is

$$f^{(t)} = \sigma(W_{fx}X^{(t)} + W_{fh}h^{(t-1)} + W_{fc}c^{(t)} + \theta_f). \quad (16)$$

The memory sequence can be updated as follows relying on input and output gates:

$$c^{(t)} = i_t * \tan h(W_{cx}X^{(t)} + W_{ch}h^{(t-1)} + \theta_c) + f^{(t)} * c^{(t-1)}. \quad (17)$$

The output gate can be used to scale output sequence, and the detailed method is as follows:

$$o^{(t)} = \sigma(W_{ox}X^{(t)} + W_{oh}h^{(t-1)} + W_{oc}c^{(t)} + \theta_o). \quad (18)$$

Finally, we can obtain

$$h^{(t)} = o^{(t)} * \tan h(c^{(t)}), \quad (19)$$

and the result can be transferred into RNN.

**2.3. BiLSTM.** Bidirectional Long Short-Term Memory (BiLSTM) [18] is a variant of traditional neural network and combination of forward LSTM and backward LSTM. Output of the model can be represented as

$$\begin{aligned} \vec{h}^{(t)} &= \overline{\text{LSTM}}(h^{(t-1)}, X^t, c^{(t-1)}), \\ \overleftarrow{h}^{(t)} &= \overline{\text{LSTM}}(h^{(t+1)}, X^t, c^{(t+1)}), \\ h^{(t)} &= \left[ \vec{h}^{(t)}, \overleftarrow{h}^{(t)} \right]. \end{aligned} \quad (20)$$

Let us take the mean of  $h^{(t)}$  as the output; that is to say, the output is mean( $h^{(t)}$ ). Until the long-short sequence has arrived at BiLSTM layer, gate structure began to carry adoption and releasing of the information through sigmoid, and the output is between 0 and 1 (1 means complete adoption, and 0 means complete discarding). The ideal structure of BiLSTM is shown in Figure 1.

**2.4. CNN.** Convolutional neural network (CNN) [18] is a feedforward neural network containing convolution operation, and its model structure generally includes input layer, convolution layer, pooling layer, full-connection layer, and output layer. The convolution layer, pooling layer, and full-connection layer can all be seen as hidden layers. Among them, the role of the convolution layer is to carry out feature extraction, and the feature extraction of input layer data can be realized by using the set filter. The specific method is shown as follows:

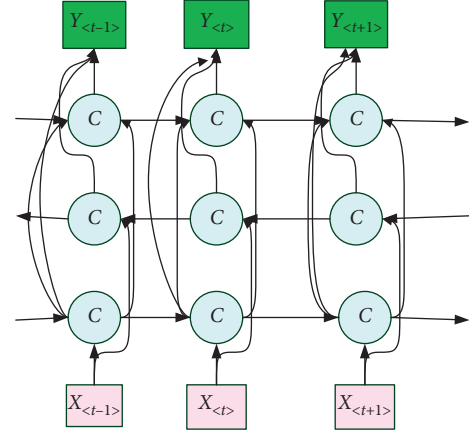


FIGURE 1: Structure of BiLSTM.

$$J_i = f(\omega * X_{i:i+g-1} + \theta). \quad (21)$$

Here,  $\omega$  denotes convolution kernel,  $g$  denotes the size of convolution kernel,  $X_{i:i+g-1}$  denotes articulatory feature vector from frame  $i$  to frame  $i+g-1$ , and  $\theta$  denotes bias value. Thus, we can obtain the feature matrix  $J = [c_1, c_2, \dots, c_{n-g+1}]$  through the convolution layer calculation.

Using max pooling technology, pooling layer can downsample the feature matrix and achieve optimal solution of the local value. Full-connection layer is located in the last layer of the hidden layer and can expand the feature diagram with topological structure to activate function. Output layer uses logical function or Softmax function to output classified label and predicted value.

### 3. Methods

**3.1. Speaker Normalization Based on Prussian Transformation.** Because speakers' articulatory characteristics are easily influenced by the speakers themselves, including their vocal tract characteristics, height, and sitting position; these factors are inherent differences between speakers. In order to eliminate these inherent differences and better quantify the kinematic characteristics of speech, we used the Prussian Transformation to normalize the articulatory characteristics of different speakers. The specific processing is shown in Figure 2.

The algorithm realized the linear geometric transformation from the original multipoint object to the target multipoint object, including scale transformation, translation transformation, and rotation transformation. It is supposed that the raw articulatory data was  $D_1$ ; then the normalization of  $D_1$  was  $D_3$ , and the target speaker's articulatory data was  $D_2$ . Using hybrid transform consisting of scale transform and rotation transform, we can take the relation between  $D_1$  and  $D_3$  as follows:

$$D_3 = HD_1a + b, \quad (22)$$

where the normalizing parameter  $\{H, a, b\}$  can be optimized relying on minimized Root Mean Square Error between target data  $D_2$  and the normalized data of raw speaker's articulatory  $D_3$ .



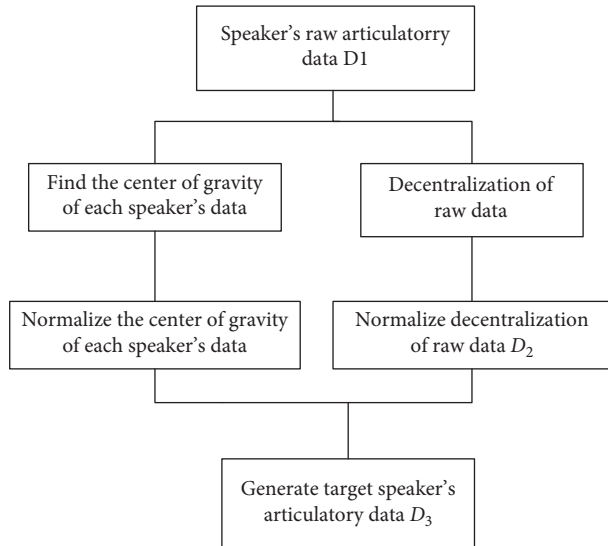


FIGURE 2: Diagram of speaker normalization algorithm based on Prussian Transformation.

To be specific, rotation matrix can be calculated using singular value decomposition, which is shown as follows:

$$(D_1)^T D_2 = U \Sigma V^T, \quad (23)$$

Here,  $\Sigma$  is the diagonal matrix,  $U$  and  $V$  are separate orthogonal matrices, and  $A$  is the diagonal matrix in which the absolute value of the diagonal elements is 1.

**3.2. BiLSTM-CNN-Based Articulatory-to-Acoustic Conversion.** According to Sections 2.2 and 2.3, CNN has a good ability to extract local features, and BiLSTM network has a good performance on the coherence of previous frames and semantic features based on word-attention mechanism [27]. This paper combined CNN and BiLSTM and used the theory of word attention to achieve articulatory-to-acoustic conversion, where BiLSTM used context information to analyze the articulatory features and train continuous frames, and word-attention layer used word-attention mechanism to extract semantic features and send them to the BiLSTM for training. In the later stage, the CNN was mainly composed of convolutional layer, pooling layer, and full-connection layer. Finally, acoustic features are output by regression layer. The specific model structure is shown in Figure 3.

As illustrated in Figure 3, the LSTM cells at each layer in the BiLSTM-CNN were divided into two parts to capture the forward and backward dependency, respectively. In this case, the forward and backward articulation feature sequences were both 10 frames and the feature vector of each frame was 8 dimensions, and the semantic feature was 1 dimension. Thus, the input feature dimension of the feature fusion layer was 169 dimensions. In the CNN part, we used 4 full-connection layers, the convolution layer with size of 128 dimensions, and the regression layer.

## 4. Experiments and Results

### 4.1. Materials

**4.1.1. Participants.** In the study, ten participants (5 males and 5 females) were recruited; all of them were aged between 25 and 40 years (average of ages is 27.1, and the STD is 1.94) with no professional language training and no orofacial surgery history [28]. Before collecting the data, all subjects were told the processes for collecting data and signed informed consent. The study was approved by the Health Sciences Research Ethics Board at Institute of Psychology of the Chinese Academy of Sciences (No. H16012).

**4.1.2. Textual Material.** Disyllable words and sentences of neural affect were chosen as textual material. Sentences of neural textual material were chosen as the spoken material, including “Xia4 yu3 le1.” (It is raining.), “Jin1 tian1 shi4 xing1 qi1 yi1.” (Today is Monday.), “Wo3 xiang3 gei3 ta1 yi2 ge4 jing1 xi3.” (I want to give him a surprise.), “Ni3 yuan2 lai2 zai4 zhe4 li3” (So you here.), “Wo3 cuo4 le1.” (I am wrong.), “Ni3 xue2 de1 zhen1 kuai4.” (You learn fast.), and “Wo3 men2 shi4 lao3 tong2 xue2.” (We are old classmates.).

Disyllable words were chosen as the spoken material, including “Mama” (Mum), “Zaijian” (Good-bye), “Tiantian” (Everyday), “Daqi” (Encourage), and “Nihao” (Hello).

**4.1.3. Data Collection.** All articulatory data and acoustic data were collected using the AG501 [29] EMA device of Carstens [29] (Lengler, Germany) as shown in Figure 4, which has 24 articulatory channels and one audio channel with 250 Hz and 48 kHz sampling rate. AG501 is widely used in electromagnetic articulography, which allows the collecting in 3D of the movements of the articulators with high precision.

We have glued 6 sensors (2 mm \* 3 mm) with thin wires to the left and right mastoids, nose bridge, and the bite plane to carry head collection and 9 sensors to the upper and lower lips, left and right lip corners, upper and lower incisors, and tongue tip, tongue mid, and tongue root (as shown in Figure 5). All subjects engaged in conversation for approximately 5 minutes after sensors were attached to provide subjects the opportunity to familiarize themselves with the presence of the sensors in the oral cavity.

The collection experiment has been carried out in the quiet environment with a maximum background noise of 50 dB. Acoustic data was collected by a match condenser microphone EM9600, and articulatory data was collected in synchronization with the acoustic data.

**4.1.4. Data Processing and Feature Extraction.** The collected data were loaded into the VisArtico, a visualization tool for filtering using a low-pass filter (cut-off is 20 Hz). Meanwhile, the articulatory data were corrected for head movement using Cs5normpos tool, which is a kind of tool in the EMA control system of AG501.

The VisArtico program can visualize kinematic data while also allowing for calculation of tongue kinematic

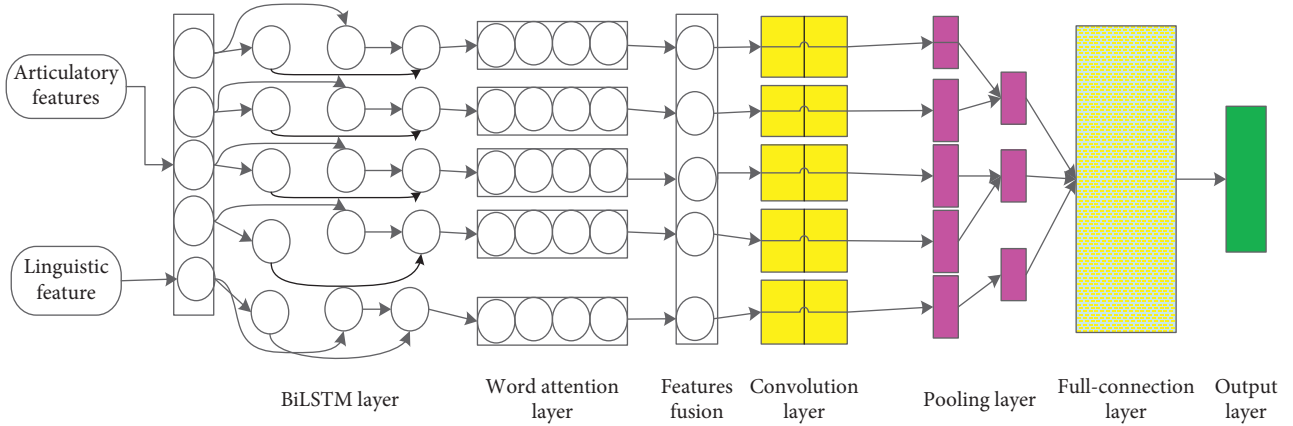


FIGURE 3: Integration model of BiLSTM-CNN.



FIGURE 4: Sensor placements on occlusal bite plane.

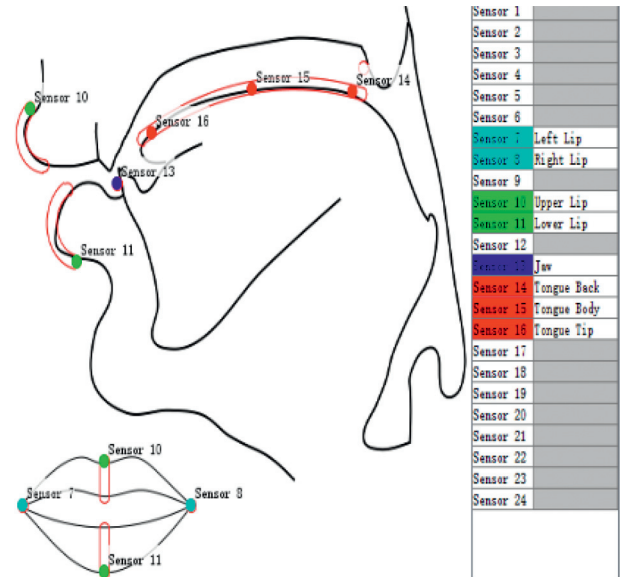


FIGURE 5: Sensor placements on the articulator.

parameters. In this paper, we extracted 8-dimensional articulatory features as shown in Table 1.

In this paper, we have chosen 299 samples of disyllables and sentences and then took 200 samples as the training data and 99 samples as the test data, respectively.

**4.2. Model Comparison of EMA-to-F2 Conversion.** In the EMA-to-F2 experiment, we compared the performances of the GMM-based, RNN-based, and BiLSTM-CNN-based methods. The Root Mean Square Error (RMSE) in Hz between the true and the predicted F2 was adopted as the evaluation measure parameter.

As a classical prediction model, GMM can approximate any function as long as the number of mixing elements is sufficient. In this study, we selected GMM with 500 Gaussian elements to accurately describe the joint probability density function of the articulatory features and acoustic features.

According to the maximum likelihood criterion, the conditional probability of acoustic features has been obtained by approximate calculation of the joint probability density function of acoustic features and articulatory features, and the closed solution of the best acoustic features has been obtained. The result is shown in Figure 6 (the figure takes 80 frames of data as the example).

For the EMA-to-F2 conversion based on BiLSTM-RNN, the 21-frame input window (10 frames forward and 10 frames backward) has been used to train the network. We have trained 50 iterations for BiLSTM-RNN with 5 hidden layers and 100 hidden units per hidden layer. The training results are shown in Figure 7, which illustrated the RMSE and loss of training data. Along with increasing the iterations number, the RMSE between the true and predicted data and loss function values decreased. The optimal model occurred at the 48th epoch, and the loss function value and RMSE reached their minimum, respectively.

The BiLSTM-CNN we proposed consisted of BiLSTM, word-attention layer, and the CNN (convolutional layer,

TABLE 1: Articulatory features.

| Abbreviation for features | Description of features                                      |
|---------------------------|--|
| Tongue root-X             | The normalized back-forward velocity of the tongue root      |
| Tongue root -Z            | The normalized up-down velocity of the tongue root           |
| Tongue body-X             | The normalized back-forward velocity of the tongue mid       |
| Tongue body-Z             | The normalized up-down velocity of the tongue mid            |
| Tongue tip-X              | The normalized back-forward velocity of the tongue tip       |
| Tongue tip-Z              | The normalized up-down velocity of the tongue tip            |
| Tongue constriction       | The kinematics range of tongue tip at back-forward direction |
| Lip aperture              | The open degree of lips at up-down direction                 |

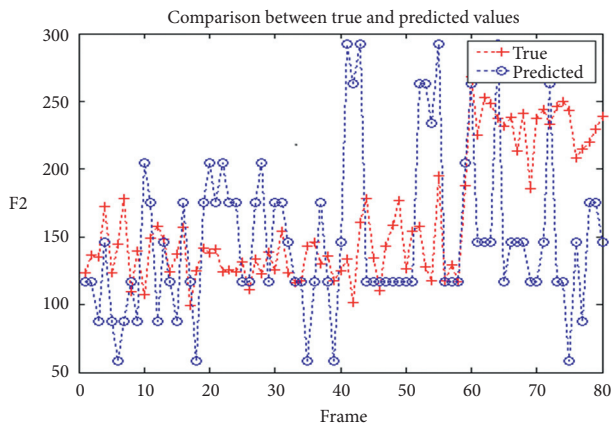


FIGURE 6: Comparison between true and conversed values of GMM-based F2.

pooling layer, full-connection layer, and regression layer). About the CNN part, we have chosen the convolutional layer with size of  $169 * 169$ , 4 full-connection layers, and the 1-dimensional regression layer. About the BiLSTM part, we took 5 hidden layers with 100 hidden units per hidden layer and adopted 21 frames (10 frames forward, 1 current frame, and 10 frames backward) as the input feature; meanwhile, the semantic feature needs input to the BiLSTM for feature fusion and training. In the training process, we initially set the learning rate to 0.005 and fixed the momentum at 0.8, with maximum epochs of 50. Then, we can find that BiLSTM-CNN is much better than the BiLSTM-RNN and GMM conversion model, and the comparisons of F2 between true value and the predicted values, using the GMM, BiLSTM-RNN, and BiLSTM-CNN based on word attention, all are shown in Figure 8.

From the figure, we can find that the predicted F2 using BiLSTM-CNN is most similar to the true value, and the predicted F2 using BiLSTM-RNN is less similar to BiLSTM-CNN. Furthermore, we used test data on GMM, BiLSTM-RNN, and BiLSTM-CNN based on word attention; the RMSE and correlation coefficient  $r$  of F2 can be obtained and shown in Table 2.

The correlation coefficient  $r$  has been used to analyze the correlation between the predicted features and the true features using Pearson product moment correlation method, which is a method to analyze the linear relationship between two variables. Here, it is supposed that there are two databases: articulatory feature input ( $x = \{x_1, x_2, \dots, x_n\}$ ) and

acoustic features output ( $y = \{y_1, y_2, \dots, y_n\}$ ), and the size of the database is  $n$ . Thus, Pearson correlation coefficient can be defined as

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (24)$$

where  $\bar{x}$  and  $\bar{y}$  represent the means of sample features  $x$  and  $y$  and  $x_i$  and  $y_i$  show  $i$ th values of  $\bar{x}$  and  $\bar{y}$ , respectively. Correlation coefficient  $r$  can reflect the strength information of the linear relationship between variable sets  $x$  and  $y$ , ranging from  $-1$  to  $1$ . If  $x_i$  and  $y_i$  are multidimensional vector, the dimensionality of the vector should be reduced first, and then the correlation analysis should be carried out.

In the study, we can find that there are strong positive correlations between the predicted and true features on all three models, which is shown in Table 1. In detail, the correlation was, in turn, BiLSTM-CNN > BiLSTM-RNN > GMM.

**4.3. Model Comparison of EMA-to-MFCC Conversion.** In the EMA-to-MFCC experiment, we adopted MMCD as the parameter to evaluate the results of articulatory-to-MFCC conversion, which can be defined as the mean value of Euclidean distance between the predicted value and true value. Here, we used 12-dimensional MFCC as the acoustic feature and compared the performances of the GMM-based, RNN-based, and BiLSTM-CNN-based methods.

In the experiment, we selected GMM with 500 Gaussian elements to accurately describe the joint probability density function of the articulatory features and acoustic features. For the BiLSTM-CNN, we set the convolutional layer with size of  $169 * 169$ , 4 full-connection layers, the 1-dimensional regression layer, and 5 hidden layers with 100 hidden units per hidden layer and adopted 21 frames (10 frames forward, 1 current frame, and 10 frames backward) as the input feature.

In the training process, we initially set the learning rate to 0.005 and fixed the momentum at 0.9, with the maximum epochs of 60. Then, we can find that BiLSTM-CNN is much better than the BiLSTM-RNN and GMM conversion model, and the comparison results are shown in Table 3.

From the table, the MMCD of BiLSTM-CNN is the minimum among three models, and BiLSTM-RNN is better than GMM but not better than BiLSTM-CNN. Meanwhile, we can find that there are strong positive correlations

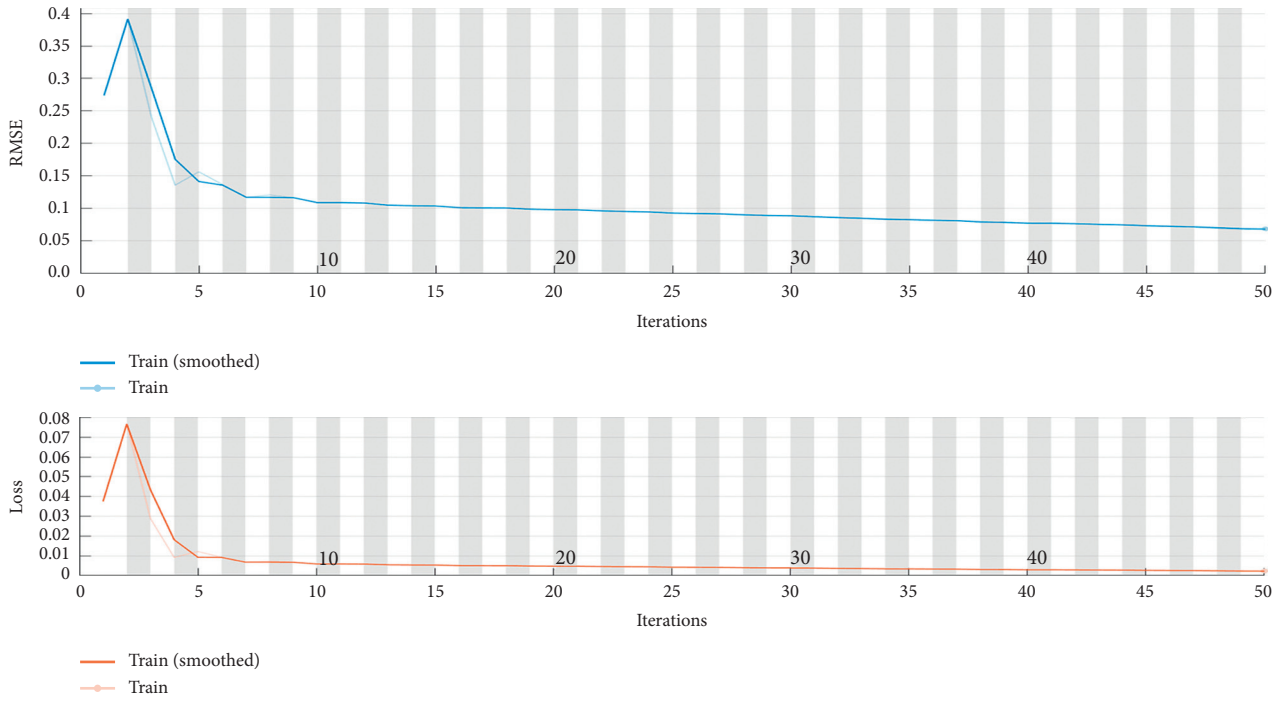


FIGURE 7: The RMSE and loss function values on the training database of BiLSTM-RNN.

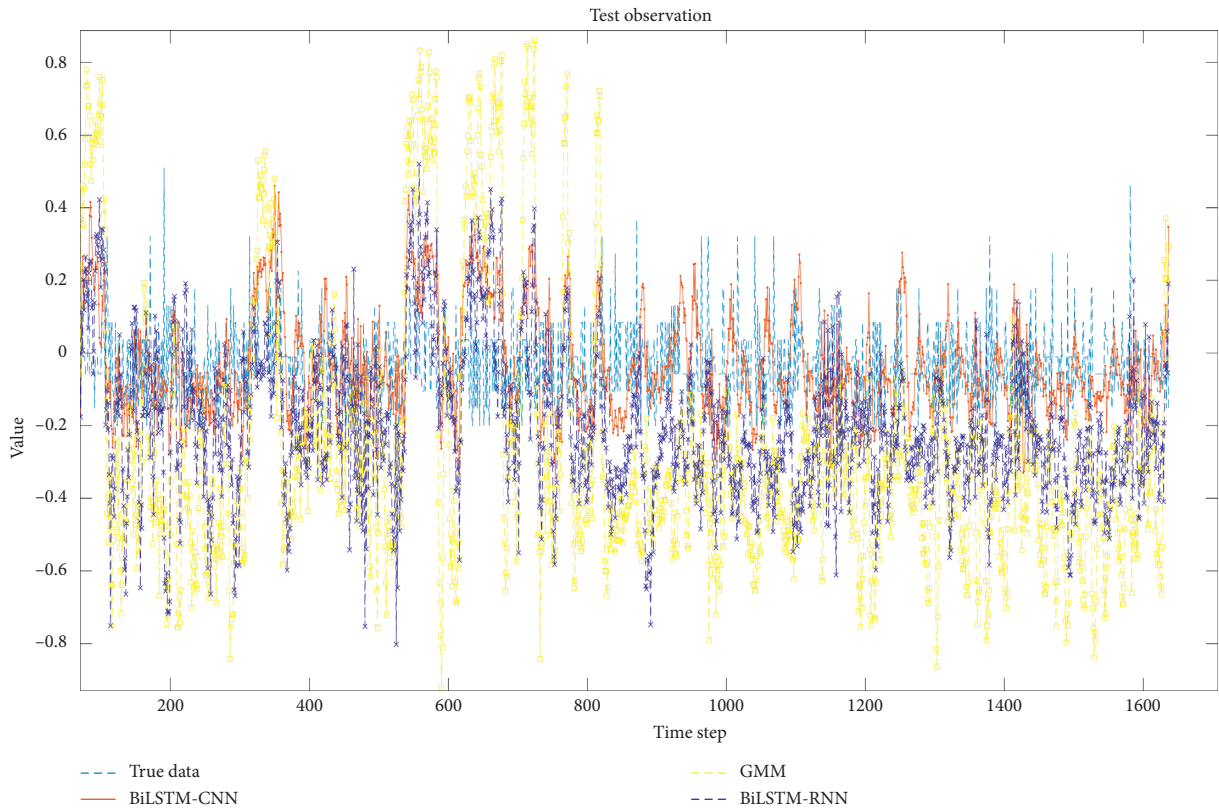


FIGURE 8: Kinematics comparison of F2 based on different network.

TABLE 2: The RMSE (Hz) and  $r$  on the test set when using GMM, BiLSTM-RNN, and BiLSTM-CNN for EMA-to-F2 conversion.

|      | GMM   | BiLSTM-RNN | BiLSTM-CNN |
|------|-------|------------|------------|
| RMSE | 37.13 | 26.42      | 22.10      |
| $r$  | 0.565 | 0.625      | 0.738      |

TABLE 3: The MMCD (dB) and  $r$  on the test set when using GMM, BiLSTM-RNN, and BiLSTM-CNN for EMA-to-MFCC conversion.

|      | GMM   | BiLSTM-RNN | BiLSTM-CNN |
|------|-------|------------|------------|
| MMCD | 2.384 | 1.824      | 1.467      |
| $r$  | 0.576 | 0.725      | 0.753      |

between the predicted and true features on all three models; in detail, the correlations are, in turn, BiLSTM-CNN > BiLSTM-RNN > GMM.

## 5. Discussion and Conclusion

This study provided a novel conversion method combining BiLSTM, CNN, and word-attention theory. In the current study, features of the tongue and the lip in 3D coordinates of AG501 have been extracted for the conversion and recognition research and acoustic features (i.e., F2 and MFCC).

From the conversion research, we found that the kinematics of tongue and lips can construct a simple graph, which are found from the application of CNN, because CNN has been used to graph signal processing widely. Meanwhile, because the database we used is Mandarin, as a kind of tone language, semantic feature plays an important role in the speech processing, especially in articulatory-to-acoustic conversion and speech recognition. So, we adopted word-attention theory in this study and achieved ideal effect, which proves that the semantic feature is helpful to the conversion study especially in Mandarin.

The current study broke the limitation of focusing on the vowel only and fused the semantic features and articulatory features. Due to the limitation of numbers of samples, we choose 299 disyllables only in this paper; the sample size was a little bit small, which will be considered in future efforts. The study in this paper should be the basement of the research of speech recognition and speech production. It can promote the fusion of artificial intelligence and Smart Campus in the future.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

Thanks are due to all the subjects in current experiment, to Guicheng Shao for technical assistance, to Jianmei Fu for

modal design, and to Jianzheng and Dong Li for assistance in data collection. This work was supported by the Educational Reform Innovation Project of Shanxi Province of China (J2019174), Science and Technology Project of Xinzhou Teachers University (2018KY15), and Academic Leader Project of Xinzhou Teachers University.

## References

- [1] R. John, "X-ray microbeam speech production database," *The Journal of the Acoustical Society of America*, vol. 88, no. S1, p. 56, 1990.
- [2] SAIL, "MRI-TIMIT: a multimodal real-time MRI articulatory corpus," 2014, <https://sail.usc.edu/span/mri-timit/ed>.
- [3] T. G. Csapó, T. Grósz, G. Gosztolya, L. Tóth, and A. Markó, "DNN-based ultrasound-to-speech conversion for a silent speech interface," in *Proceedings of the 2017 INTERSPEECH*, Stockholm, Sweden, 2017.
- [4] Y. Luo, *A Study on the Location and Coarticulation of Consonants Based on EPG—A Case Study of Zhuang and Miao Languages*, Master, *Experimental Phonetics*, Shang Normal University, Shanghai, China, 2017.
- [5] K. Richmond, "Preliminary inversion mapping results with a new EMA corpus," in *Proceedings of the 2009 INTERSPEECH*, Brighton, UK, 2009.
- [6] K. Richmond, Z. Ling, J. Yamagishi, and B. Ur, "On the evaluation of inversion mapping performance in the acoustic domain," in *proceeding of the 2013 INTERSPEECH*, Lyon, France, 2013.
- [7] S. Ouni and Y. Laprie, "Modeling the articulatory space using a hypercube codebook for acoustic-to-articulatory inversion," *The Journal of the Acoustical Society of America*, vol. 118, no. 1, pp. 444–460, 2005.
- [8] S. King and A. Wrench, "Dynamical system modeling of articulator movement," in *Proceedings of the 1999 ICPHS*, San Francisco, CA, USA, 1999.
- [9] S. Dusan and L. Deng, "Acoustic-to-articulatory inversion using dynamical and phonological constraints," in *Proceedings of the 2000 Seminar on Speech Production*, Sydney, Australia, 2000.
- [10] Z. L. Korin Richmond and J. Yamagishi, "Benigno UR and IA, on the evaluation of inversion mapping performance in the acoustic domain," in *Proceedings of the 2013 INTERSPEECH*, Lyon, France, 2013.
- [11] T. Toda, A. W. Black, and K. Tokuda, "Acoustic-to-articulatory inversion mapping with gaussian mixture model," in *Proceedings of the 2004 INTERSPEECH*, Jeju Island, Republic of Korea, 2004.
- [12] S. Hiroya and M. Honda, "Estimation of articulatory movements from speech acoustics using an HMM-based speech production model," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 2, pp. 175–185, 2004.
- [13] J. Lin, W. Li, Y. Gao et al., "Improving Mandarin tone recognition based on DNN by combining acoustic and articulatory features using extended recognition networks," *Journal of Signal Processing Systems*, vol. 90, pp. 1077–1087, 2018.
- [14] Z.-H. Ling, K. Richmond, and J. Yamagishi, "An analysis of HMM-based prediction of articulatory movements," *Speech Communication*, vol. 52, no. 10, pp. 834–846, 2010.
- [15] L. Badino, C. Canevari, L. Fadiga, and G. Metta, "Integrating articulatory data in deep neural network-based acoustic modeling," *Computer Speech & Language*, vol. 36, pp. 173–195, 2016.

- [16] C. C. Leonardo Badino, L. Fadiga, and G. Metta, "Deep-level acoustic-toarticulatory mapping for DBN-HMM based phone," in *Proceedings of the 2012 IEEE Spoken Language Technology Workshop (SLT)*, Miami, FL, USA, 2012.
- [17] J. Bai, F. Li, and H.-D. Ji, "Attention-based BiLSTM CNN Chinese microblog position detection model," *Computer Applications and Software*, vol. 3, no. 35, pp. 266–274, 2018.
- [18] K. Sun, "Word attention-based BiLSTM and CNN ensemble for Chinese sentiment analysis," *Computer Science and Application*, vol. 10, no. 2, pp. 312–324, 2020.
- [19] Z.-C. Liu, Z.-H. Ling, and L.-R. Dai, "Articulatory-to-acoustic conversion with cascaded prediction of spectral and excitation features using neural networks," in *Proceedings of the 2016 INTERSPEECH*, San Francisco, CA, USA, 2016.
- [20] Z.-C. Liu, Z.-H. Ling, and L.-R. Dai, "Articulatory-to-acoustic conversion using BLSTM-RNNs with augmented input representation," *Speech Communication*, vol. 99, pp. 161–172, 2018.
- [21] Y.-M. Wang and K. Chen, "End-to-end audio-visual dual-mode speech recognition based on SDBN and BLSTM attention fusion," *Communication Science*, vol. 12, pp. 80–90, 2019.
- [22] L.-Y. Wang, C.-H. Liu, D.-B. Cai et al., "Text emotion analysis based on CNN-BiLSTM network with attentional model," *Journal of Wuhan University of Technology*, vol. 4, no. 41, pp. 387–394, 2019.
- [23] B. An Ji, *MSEE, Speaker Independent Acoustic-to-Articulatory Inversion*, Doctor, Electrical and Computer Engineering, Marquette University, Milwaukee, WI, USA, 2015.
- [24] A. Ji, J. J. Berry, and M. T. Johnson, "Vowel production in Mandarin accented English and American English: kinematic and acoustic data from the Marquette University Mandarin accented English corpus," *Speech Communication*, vol. 19, Article ID 060221, 2013.
- [25] R. Li, Z. Wu, Y. Huang, J. Jia, H. Meng, and L. Cai, "Emphatic speech generation with conditioned input layer and bidirection LSTMs for expressive speech synthesis," in *Proceedings of the 2018 ICASSP*, Calgary, Canada, 2018.
- [26] J. C. Aviles and A. Kouki, "Position-aided mm-wave beam training under NLOS conditions," *IEEE Access*, vol. 4, pp. 8703–8714, 2016.
- [27] L. Wu, F. Tian, L. Zhao, J. Lai, and T.-Y. Liu, "Word attention for sequence to sequence text understanding," in *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, pp. 1–8, New Orleans, LA, USA, 2018.
- [28] G.-F. Ren, X.-Y. Zhang, D. Li, and etal, "Design and evaluation of Mandarin bi-modal emotion speech database," *Modern Electronic Technology*, vol. 41, no. 14, pp. 182–186, 2018.
- [29] M. Stella, A. Stella, F. Sigona, P. Bernardini, M. Grimaldi, and B. G. Fivela, "Electromagnetic articulography with AG500 and AG501," in *Proceedings of the 2013 INTERSPEECH*, Lyon, France, 2013.

## Research Article

# Optimization and Simulation of a Reasonable Scheduling Model under Multiple Tasks in Company Management

Zhuo Wang <sup>1,2</sup>

<sup>1</sup>School of Economics and Management, Yanshan University, Qinhuangdao 066000, Hebei, China

<sup>2</sup>School of Management, Jilin Normal University, Siping, Jilin 136000, China

Correspondence should be addressed to Zhuo Wang; wangzhuo@jlnu.edu.cn

Received 20 July 2020; Revised 29 August 2020; Accepted 12 September 2020; Published 25 September 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Zhuo Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of market economy, the task scheduling model has become the core problem in the field of corporate management. In order to solve the problem that the stability of the model will decline due to the interference of human factors in the process of multitask scheduling in the traditional algorithm, a reasonable scheduling model based on the priority principle is proposed in this paper. This paper expounds the principle of multitask scheduling in company management, constructs a multitask scheduling network model based on priority, performs virtual scheduling on the model, searches the optimal solution in the solution space of the scheduling problem, and obtains the reasonable multitask scheduling method in the company management. Through the analysis of relevant simulation experiments, it can be concluded that the task scheduling algorithm proposed in this paper can not only allocate the corresponding resources for high-priority tasks effectively, but also can save the cost, so as to show better execution efficiency.

## 1. Introduction

With the rapid development of market economy, the competition between enterprises is becoming more and more fierce. As a more important factor affecting the competitiveness of enterprises, the method of task scheduling in company management has been paid more and more attention by company managers [1, 2]. Therefore, the method of task scheduling in company management has become a hot topic in the field of enterprise, which has been paid attention by many experts. At present, the main task scheduling models in company management mainly include the task scheduling method based on the fuzzy clustering algorithm, the task scheduling method based on the optimization genetic algorithm, and the task scheduling method based on the ant colony algorithm [3, 4]. Because of the wide development space of the task scheduling method, it has become the focus of many experts.

Task scheduling model has become the core problem in the field of company management. In the management of the company, the pipeline work task management mode is

adopted, and the multithread and multitask mode is adopted for the work task under this management [5]. In the process of multitask scheduling, the correlation between tasks is very complex, which is greatly affected by the scheduling order. The traditional task scheduling model is based on single linear order optimization [6, 7]. Once there are too many tasks, the contradiction between different task scheduling will increase, the stability of the model will decline, and the efficiency of scheduling will be low. However, it is difficult to assign a schedulable priority to each task in the design model due to the limitation of multiple times. At the same time, it is difficult to avoid the undesirable priority inversion. Therefore, the design model cannot be used to analyze the fixed priority scheduling theory.

Aiming at the defects of the traditional algorithm, a reasonable scheduling method based on the multiconstraint scheduling model is proposed. In this paper, the principle of multitask scheduling in company management is described in detail [8, 9]. A multitask scheduling network model based on priority is constructed by using this principle. Virtual scheduling is carried out for the network model, and the

optimal solution is searched in the solution space, and the reasonable multitask scheduling method in company management is obtained [10]. The experimental results show that the improved algorithm can improve the resource utilization rate and scheduling efficiency and avoid the defect of reduced stability of the scheduling model caused by multiwork doping.

## 2. Principle of Multitask Scheduling in Company Management

*2.1. Analysis of Task Scheduling Model.* Task scheduling is simply a problem of allocating resources according to time to complete tasks. The accuracy and rationality of scheduling is directly related to the economic effect of the whole production activity. The goal of task scheduling software is to use advanced computer technology to achieve efficient control of production scheduling [11]. The objectives of the task scheduling software are as follows:

- (1) *Convenient Interaction.* The software should have a friendly man-machine interface, which is convenient, intuitive, and easy to learn. The user can easily realize the man-machine dialogue.
- (2) *Network Management.* As a part of the enterprise internal management system, the task scheduling software should have a good information interface to realize the information interaction with other sub-systems. Through the interface, it can realize the functions of task allocation, release, and processing.
- (3) *Controllability of Scheduling Process.* It has the monitoring function for the whole process of task scheduling and timely reflects the problems to the system administrator.
- (4) *Security.* Most of the internal resources involved in software are confidential, so security is very important. The software should have the ability of user identification and authority control.

According to the requirement analysis of task scheduling software, the task scheduling is divided into six functions: plan preprocessing, task scheduling, basic information management, material resource management, job management, and work process tracking. Thus, the domain model of task scheduling is obtained, as shown in Figure 1. The task scheduling model is mainly composed of two modules: task request scheduling and optimal task selection. The classifier classifies tasks into several classes according to their priority and sends them to the corresponding queue buffer and gives each queue a weight. The difference between virtual scheduling and actual scheduling is that the virtual scheduling task allocates a set of resources, while the actual scheduling task allocates a specific resource. Virtual scheduling is developed based on relaxation theory, so virtual scheduling is an intermediate process to solve the scheduling scheme.

After the plan preprocessing function searches the plans, it classifies the plans according to the retrieved results and compiles the dispatch list, which is used to

drive the task scheduling [12]. After receiving the dispatch list issued by the plan, the task scheduling first calls the residual production capacity information in the basic information management, the process information of the product, and the task decomposition, assignment, and sorting based on the example-based contract net scheduling method. Then, the raw materials and semifinished products in the resource management are called to schedule and process the tasks [13, 14]. After the scheduling is completed, the production, man hour quota, and cost are statistically managed. The whole scheduling process is accompanied by work process tracking. Work process tracking tracks the execution status of the plan before the execution of the task scheduling, tracks the logistics and cost information generated in the scheduling process, and tracks the output and quality in the operation management after the scheduling is completed.

*2.2. Design of Task Scheduling Architecture.* Class model is the core of object-oriented modeling and an important part of UML. Its main function is to define the classes in the system and the relations between them. Class structure in the class model is formed by abstracting meaningful concepts from domain requirements and then encapsulating certain behaviors with them [15, 16]. The initial class model is a concept-based class model, which focuses on the expression of the relationship between classes, while ignoring the expression of the attributes and behaviors in the internal structure of the class. According to the domain requirements and basic framework of the system, the initial class model is constructed abstractly. The following is an analysis and summary of the abstract modeling methods of the initial class model [17, 18]:

- (1) The initial class is extracted from the underlying use case, activity diagram, and information carrier analysis, and the initial class framework is established
- (2) From the use case diagram and the activity diagram to the relationship between information flow and data flow abstract classes in the class diagram
- (3) The mapping from the use case diagram and the activity diagram to workflow in the class diagram is used to reflect business processing logic
- (4) Accurate identification of association types (association, aggregation, composition, and generalization)
- (5) An interface definition or description that expresses the services (operations) provided to the user using the interface
- (6) Proxy object is used to refine association and provide access control; action object is used to provide preprocessing to simplify interaction

The initial class architecture of task scheduling software is shown in Figure 2. The initial class framework of task scheduling is developed according to the business main line of domain application, which embodies the functional abstraction and data processing flow of IPO control.



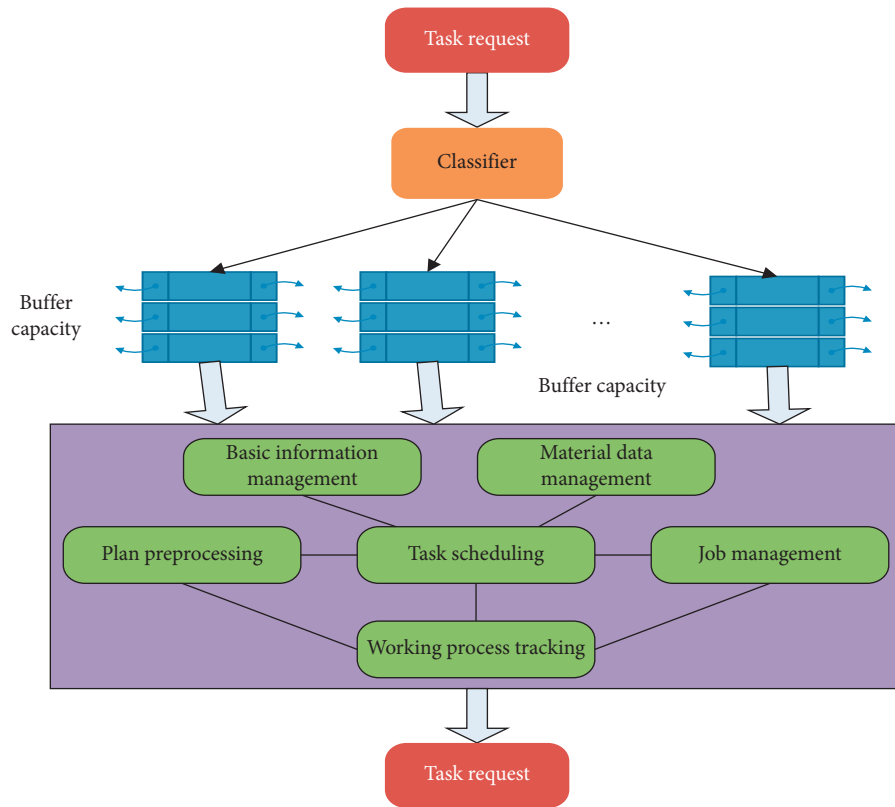


FIGURE 1: Domain model of production task scheduling.

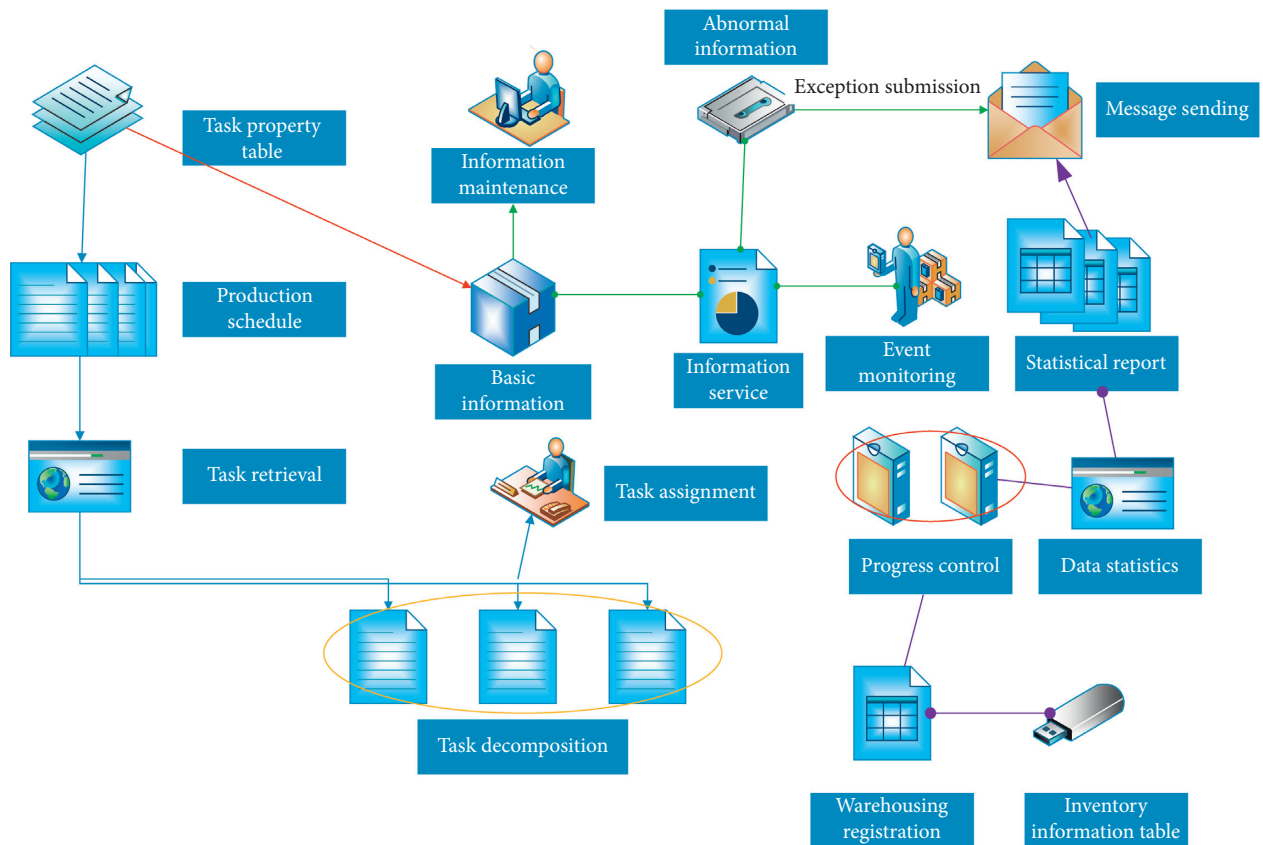


FIGURE 2: Initial class architecture of production task scheduling.

In addition, the initial architecture of the system only describes the status of this class in the system and the relationship between this class and other classes in the system, but fails to describe its detailed design and optimization. Therefore, it is necessary to further analyze and design the middle classification of the initial class framework according to the situation. The optimization method is used to optimize the structure [19–21]. Task assignment class is the core process of collaborative task scheduling. According to the domain analysis, the task assignment class is further analyzed and designed. The structure of the task assignment class is shown in Figure 3.

*2.3. Description of Multitask Scheduling Principle.* According to the above method, we can solve the utilization ratio of scheduling resources in multitasks. In the process of scheduling tasks, the demand for nonrenewable resources is determined. Therefore, the task scheduling cannot be ignored. Therefore, when solving the multitask scheduling problem, we first simplify the problem [22, 23]:

- (1) Nonupdatable resources are ignored during the operation
- (2) The duration of task completion is only related to the working capacity of the selected renewable resources

The solution of the scheduling problem with multiple tasks is  $(W, R, S)$ .  $W$  is used to describe the total number of tasks, and  $\omega_i$  is used to describe  $i$ .  $R$  is used to describe the resource allocation vector of all tasks, and  $r_i$  is used to describe the allocation of resources to task  $i$ .  $S$  is used to describe the start time of work task, and  $S_i$  is used to describe the start time of work task  $i$ .  $S$  is obtained by calculating  $W$  and  $R$ . Therefore, the process of solving the scheduling problem is to solve the  $W$  and  $R$  values. The process of solving the multitask scheduling problem can be described as follows: (1) constructing the priority network of the problem to be scheduled; (2) through virtual scheduling, the solution space is continuously compressed until the end condition is met, and the problem solving process is exited; (3) the solution result (no solution or optimal solution) of the scheduling problem is obtained.

In the management of the company, the pipeline work task management mode is adopted, and the multithread and multitask mode is adopted for the work task under this management. In the process of multitask scheduling, the correlation between tasks is very complex, which is greatly affected by the scheduling order. In the company management, reasonable distribution of work tasks is conducive to improving task efficiency. The principle of scheduling is as follows.

Set  $t = \{T_i\} (1 \leq i \leq l)$  as the multitask set in company management, which can be described by the following matrix:

$$t = \begin{bmatrix} T_1^{\text{Type}} & T_1^L & T_1^{\text{In}} & T_1^{\text{Out}} & T_1^E \\ T_2^{\text{Type}} & T_2^L & T_2^{\text{In}} & T_2^{\text{Out}} & T_2^E \\ & & \dots & & \\ T_l^{\text{Type}} & T_l^L & T_l^{\text{In}} & T_l^{\text{Out}} & T_l^E \end{bmatrix}. \quad (1)$$

The row vector of the matrix can be used to describe the attribute set of the corresponding task. Among them,  $T_i^{\text{Type}}$  is used to describe the types of work tasks;  $T_1^L$  is the total length of work tasks;  $T_i^{\text{In}}$  is the input size of work tasks;  $T_i^{\text{Out}}$  is the output size of work tasks;  $T_i^E$  is the expected value of work tasks by company management. QoS can also be divided into different tasks according to the type of QoS. For example, completion time, labor quantity, and cost.

Set  $d = \{D_i\} (1 \leq i \leq m)$  to describe the resource set of a work task in the company management, which can be described by the following matrix:

$$d = \begin{bmatrix} D_1^{\text{rg}} & D_1^{\text{time}} & D_1^{\text{sm}} & D_1^{\text{cb}} & D_1^{\text{jq}} \\ D_2^{\text{rg}} & D_2^{\text{time}} & D_2^{\text{sm}} & D_2^{\text{cb}} & D_2^{\text{jq}} \\ & & \dots & & \\ D_m^{\text{rg}} & D_m^{\text{time}} & D_m^{\text{sm}} & D_m^{\text{cb}} & D_m^{\text{jq}} \end{bmatrix}. \quad (2)$$

The row vector of the matrix is used to describe the attributes of various resources in the corresponding task. Among them,  $D_i^{\text{rg}}$  is used to describe the working ability of workers;  $D_i^{\text{time}}$  is used to describe the completion time of tasks;  $D_i^{\text{sm}}$  is used to describe the number of workers required to complete tasks;  $D_i^{\text{cb}}$  is used to describe the use cost of a single worker;  $D_i^{\text{jq}}$  is used to describe the use cost of machines. In order to comprehensively consider the performance of various resources in multitask scheduling, it is necessary to normalize the performance attribute values of resources.

$$V = \begin{cases} \frac{V_i^{\text{rg}} - \min(\text{rg})}{\max(\text{rg}) - \min(\text{rg})}, & \max(\text{rg}) \neq \min(\text{rg}), \quad 1 \leq i \leq m, \\ 1, & \max(\text{rg}) = \min(\text{rg}), \end{cases} \quad (3)$$

where  $V_i^{\text{rg}}$  is the working ability attribute value of the  $i$ th worker,  $\min(\text{rg})$  is the minimum value of the working ability performance attribute value of  $m$  workers in the scheduling resource set, and  $\max(\text{rg})$  is the maximum value of the working ability attribute value of  $m$  workers. Obviously, formula (1) normalizes the attribute value of workers' working ability to the interval of  $[0, 1]$ .

The utilization rate of scheduling the resource set under multitasks can be described by a vector, where  $D_{i,\text{uti}} (1 \leq i \leq m)$  is used to describe the corresponding scheduling resource utilization. For example, for labor-intensive tasks, the utilization rate of scheduling resources is mainly reflected by worker resources, and its calculation formula can be described as follows:

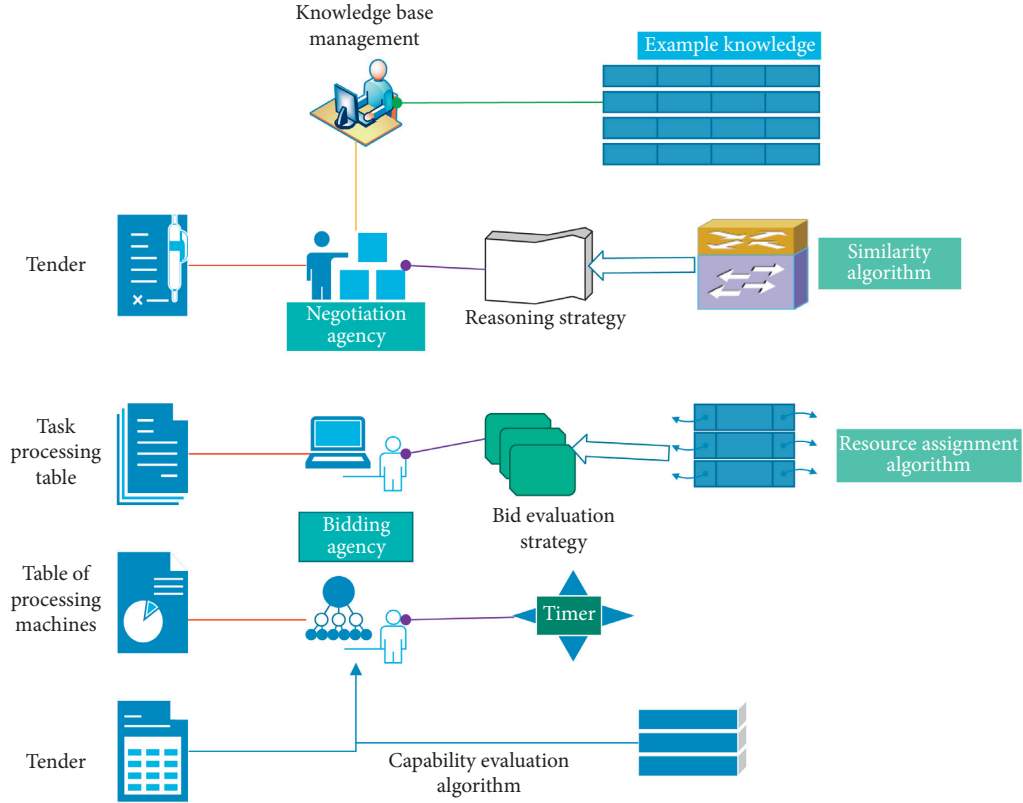


FIGURE 3: Task assignment class architecture.

$$D_{i,uti} = \frac{\sum_{j=1}^l (T_{j,etime} - T_{j,stime}) \times \Delta_{ij}}{T_{time}}, \quad 1 \leq i \leq m, \quad (4)$$

where  $T_{j,etime}$  and  $T_{j,stime}$  are used to describe the end time and the start time of the corresponding task;  $\Delta_{ij} = 1$  is used to describe that task  $J$  is allocated on schedule  $I$ , otherwise  $\Delta_{ij} = 0$ .

The overall load balance  $V_m$  of scheduling resources under multiple tasks can be described by the standard value of variance. The specific calculation formula is as follows:

$$\begin{cases} V_m = \sqrt{\frac{1}{m} \sum_{i=1}^m (V_{i,uti} - \bar{V})^2}, \\ \bar{V} = \frac{1}{m} \sum_{i=1}^m V_{i,uti}. \end{cases} \quad (5)$$

The optimization goal of scheduling is to achieve the minimum task scheduling span  $T_{j,etime}$  and the minimum system resource overall load  $V_m$  through the reasonable scheduling of task set  $t$  on the scheduling resource set  $v$ . In the management of the company, the pipeline work task management mode is adopted, and the multithread and multitask mode is adopted for the work task under this management. In the process of multitask scheduling, the correlation between tasks is very complex, which is greatly affected by the scheduling order. The traditional task

scheduling model is based on single linear order optimization. Once there are too many tasks, the contradiction between different task scheduling will increase, the stability of the model will decline, and the efficiency of scheduling will be low.

In addition, due to the need to maintain the stack of each task, there is also a lot of memory overhead. In order to reduce the space-time cost of real-time multitasking, it is common to map multiple tasks to the same thread, so that the thread will receive and process the requests of each task in the queue in turn. However, it is difficult to assign a schedulable priority to each task in the design model due to the limitation of multiple times. At the same time, it is difficult to avoid the undesirable priority inversion. Therefore, the design model cannot be used to analyze the fixed priority scheduling theory. The other method is to assign an implementation thread for each task; that is, the number of tasks is equal to the number of threads, which will have the disadvantage of high space-time cost.

### 3. Research on Multitask Scheduling Algorithm Based on Priority

**3.1. Priority Workflow Scheduling Prototype.** The level of a workflow can be determined by the level of user payment, as shown in Figure 4. For example, if a cloud service provider divides users into free users, VIP, the priority value can be 0, 1, and 2. If there are multiple workflows applying for one cloud resource at the same time, the higher the level of

workflow, the easier it is to obtain the cloud computing resources needed. If a workflow does not get the cloud computing resources it needs for a long time due to its low level, we can set a value for the workflow. Counting from 0 and adding one after each scheduling. If the value exceeds the set value after multiple rounds of scheduling, the workflow can obtain the cloud computing resources it needs regardless of its priority level in the next round of scheduling. The set value can be calculated according to the following steps.

Firstly, suppose that there are  $M$  workflows applying for the required cloud computing resources and the cloud resource provider divides the user level of cloud computing resources into  $N$  levels, then the value can be set as  $M/N$ . For the workflow being executed, its priority may be lower than that of the workflow that is applying for cloud computing resources. We prefer to let the workflow continue to execute, rather than stop its execution forcibly, because this will waste the consumed resources and make the total consumption more.

**3.2. Scheduling Algorithm Based on Workflow Priority (PISA).** PISA algorithm is based on the access control model; that is, the workflow task can enter the scheduling sequence and calculate the priority factor only after it has been authenticated by the access control model. Therefore, in PISA, this paper first introduces the access policy audit step. This is based on access control model authentication. When a workflow applies for cloud computing resources, the cloud computing resource provider first queries the access policy library and then queries the corresponding access policy according to the application information of the workflow. If the access policy is accessible, the workflow can enter the second step, that is, entering the workflow scaling sequence. First of all, the workflow task of computing resource application is authenticated. Only the task that meets the rules can enter the second step. In order to facilitate the experiment, we call the access policy of the workflow task as WAPC. The definition of WAPC is as follows:

$$\text{WAPC}[\text{work, task, } (T_s, T_c, P)], \quad (6)$$

where  $T_s$  is the start time of task  $T_i$ ,  $T_c$  is the end time of task  $T_i$ , and  $P$  is a periodic expression. The value priority defines the corresponding level of cloud computing resources that the workflow task can obtain. If the level of resource requested by the workflow task matches the value of priority, the request of the workflow will be passed. If the level of the workflow does not meet the corresponding requirements or the time quota is insufficient, the application is rejected. We define the time quota  $T_q$ , which indicates the maximum execution time quota of the task. Work defines the set priority value of the workflow to which the workflow task belongs. The authorization rules for the policy are as follows.

Suppose that user  $U$  applies to the cloud computing service provider for a workflow instance  $W$ , and the task ID contained in the request obtains the required cloud computing resources at time  $T$ . In the workflow task, the priority

value of the corresponding level resources required is priority. If  $T_s < t < T_c$ , the access application is passed, otherwise the application is rejected.

For a workflow with multiple tasks, the completion of one task may depend on the completion degree of other tasks, so their running time is related. In order to facilitate the statistics of the possible time completion rate of a task, we define the time parameter  $T_q$ , which takes the time parameters of each task into account:

$$\text{time} = \frac{\text{Min}(t_i) + \text{WS}(t_i)(t_e - t_s)}{\sum_{i=1}^m t_i}, \quad (7)$$

where  $\text{Min}(t_i)$  is the minimum execution time of task  $t_i$ ,  $\text{WS}(t_i)$  is the time parameter required for the task to use all available cloud computing resources, and  $t_q$  cannot be greater than  $t_e - t_s$ , otherwise its request will not be passed.

For a workflow  $W_i$ , it may contain multiple tasks. Similarly, users can specify the hierarchical weights of these tasks. Suppose that there are  $n$  tasks whose rank weights are  $T_1, T_2, \dots, T_n$  and  $W_i$  are the priority weights of the workflow to which these tasks belong, then the average rank weight bps of the tasks is

$$\text{bps} = \sum_{i=1}^m \frac{t_i}{N}. \quad (8)$$

Figure 5 only shows the work scheduling process, so there is no numerical value in the graph. It is only a reference for the process, and there may be other types. Figure 5 illustrates the scheduling process of workflow tasks.

For task scheduling, we also have the following considerations:

- (1) The amount of cloud computing resources that a cloud computing service provider can provide is limited.
- (2) Tasks belonging to high-level workflow should be executed first, which means that the first-level impact factor  $A$  is more important than the second-level impact factor  $B$ . In some cases, some tasks of low-level workflow may be higher in the scheduling queue after being calculated by the above formula, so we need to add the above conditions.

The second step is to bind the tasks that have been prioritized to the virtual machine and establish the mapping relationship between tasks and virtual machines. This involves the concept of quality of service. In cloud computing, service quality is used to measure the user satisfaction with cloud computing services. Different users have different standards for QoS, some may need to ensure real-time performance, while others want low cost. According to the importance of the user's task, the priority factor of the related task is calculated, and all the tasks participating in the scheduling are prioritized. Then the virtual machine resources are sorted. The higher the level of the task, the more abundant the resources can be obtained.

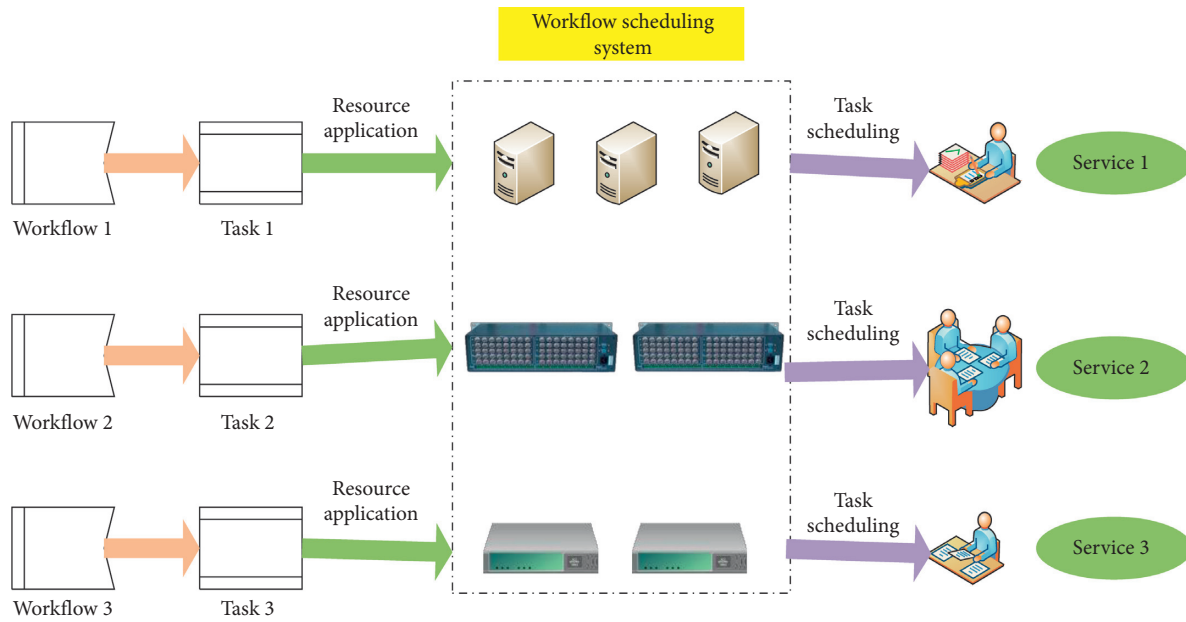


FIGURE 4: Workflow scheduling system.

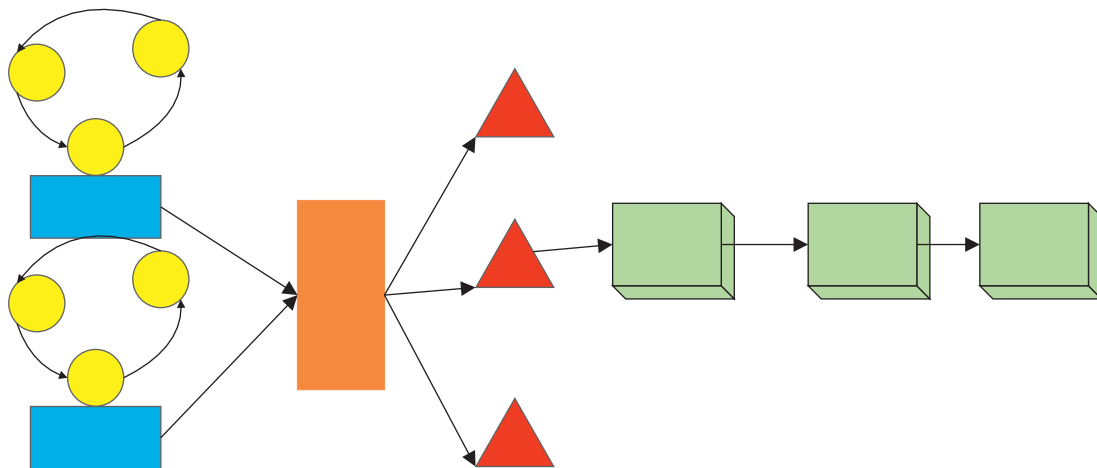


FIGURE 5: Task scheduling block diagram.

### 4. Simulation Experiment and Analysis

4.1. *GABP Optimized RO*. The cloud simulation environment used in this paper includes 10 virtual machines, each of which performs a task. We set up five users to apply to cloud computing service providers. Each user proposes a workflow, and each workflow package contains two tasks. Based on the convenience of experiments, this paper assumes that all tasks for priority factor calculation have passed the authentication of the access control model, and we only consider the priority of tasks without considering the order of task execution. In order to verify the rationality of the multitask scheduling method based on the multiconstraint scheduling algorithm, an experiment is needed. In the process of the experiment, the resource utilization, personnel utilization, and scheduling efficiency are used to

measure the performance of the multitask scheduling method. The relationship among enterprise personnel, resources, and work tasks is described in Figure 6.

The data related to the relationship among personnel, resources, and work tasks are described in Table 1.

The traditional algorithm and the improved algorithm are, respectively, used for multitask scheduling in company management. In the process of the above experiment, the personnel utilization rate of different algorithms is shown in Figure 7.

The traditional algorithm and the improved algorithm are used to schedule multitasks in company management, respectively. The scheduling efficiency of different algorithms is shown in Figure 8.

According to Table 1, it can be seen that using the improved algorithm for reasonable scheduling under

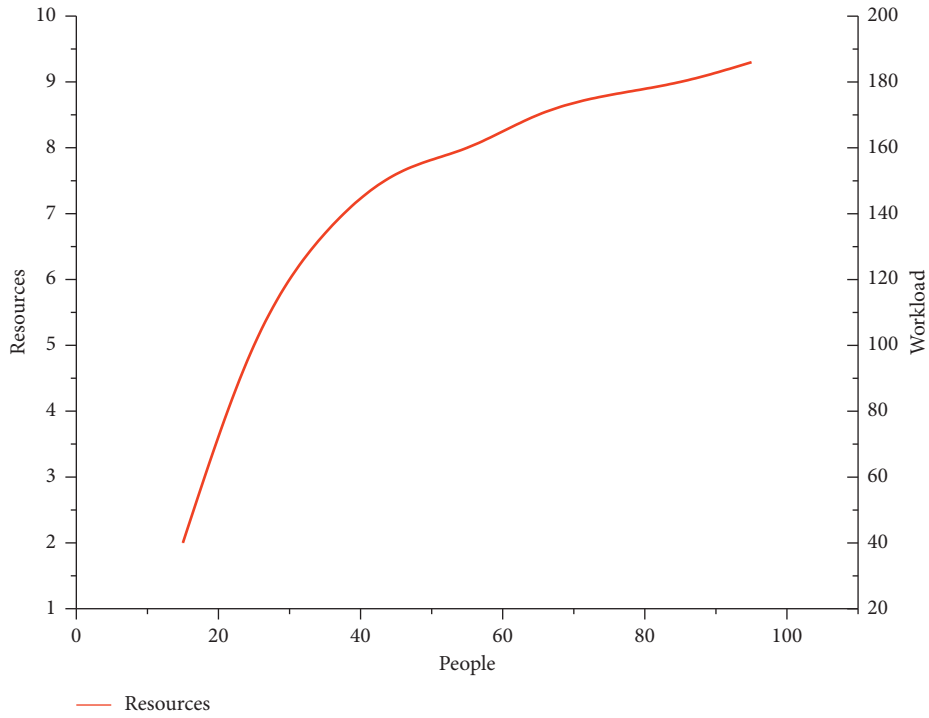


FIGURE 6: Relationship among personnel, resources, and tasks in enterprises.

TABLE 1: Comparison of resource utilization of different algorithms.

| Number of experiments | Number of personnel | Number of resources | Workload |
|-----------------------|---------------------|---------------------|----------|
| 1                     | 15                  | 2                   | 20       |
| 2                     | 25                  | 4                   | 35       |
| 3                     | 35                  | 6                   | 55       |
| 4                     | 45                  | 8                   | 70       |
| 5                     | 55                  | 10                  | 80       |
| 6                     | 65                  | 12                  | 120      |
| 7                     | 75                  | 14                  | 150      |
| 8                     | 85                  | 16                  | 180      |
| 9                     | 95                  | 18                  | 210      |
| 10                    | 105                 | 20                  | 220      |

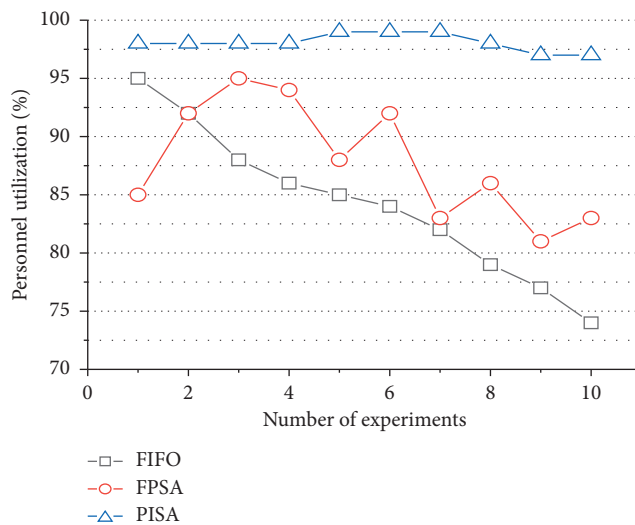


FIGURE 7: Comparison of personnel utilization of different algorithms.

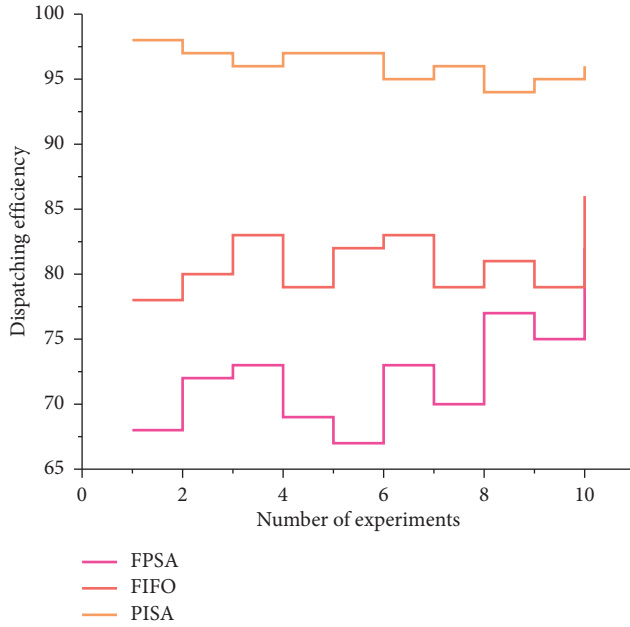


FIGURE 8: Comparison of scheduling efficiency of different algorithms.

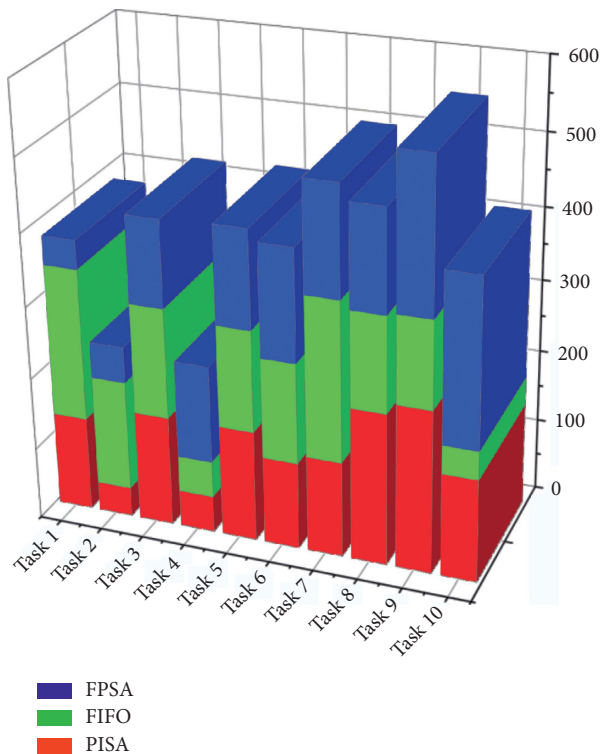


FIGURE 9: Comparison chart of execution time experiment results.

multiple tasks in company management can improve resource utilization and personnel utilization and effectively improve scheduling efficiency. The experimental data are compared and analyzed, as shown in Figures 9 and 10. Generally speaking, the task scheduling algorithm based on PISA has higher execution efficiency for high-priority tasks than those based on FIFO and fixed-priority FPSA.

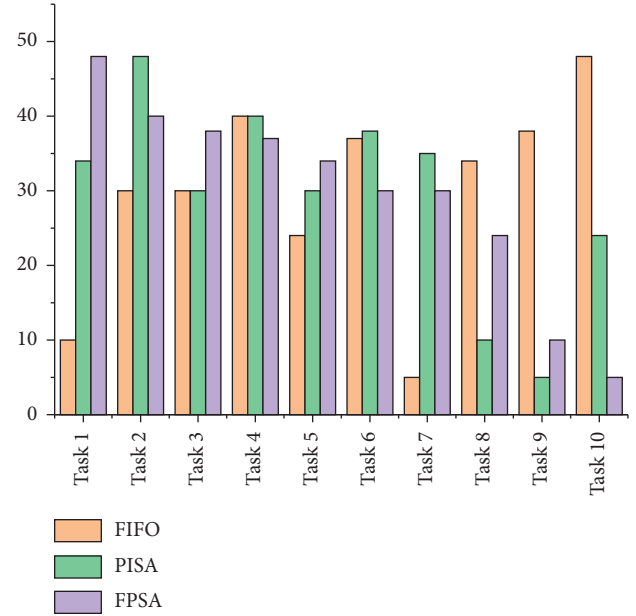


FIGURE 10: Execution cost comparison chart.

In terms of execution time, taking MS as the unit and 135 ms as the boundary, we can see in Figure 9 that in the PISA test, 5 tasks were completed within 135 ms, including 3 tasks with high priority. In the FIFO test, 5 tasks were completed in time, but only 1 was of high priority. Obviously, the task scheduling strategy based on the PISA algorithm is more suitable for the resource allocation of high-priority tasks. Compared with FPSA fixed priority, both of them can provide better services for high-priority tasks. Therefore, the efficiency difference is relatively small, and the efficiency of FPSA is higher than that of PISA because FPSA takes the maximum satisfaction of high-priority tasks as the optimization condition, and optimization is more extreme.

From the perspective of the execution cost, we can see from Figure 10 that although the fixed priority can meet the user's execution requirements for important priorities to a certain extent, its priority is fixed. Only the priority level of the workflow is considered, and the priority value of each task included in the workflow is not considered. Therefore, the implementation cost of FPSA is not as reasonable as that of PISA. PISA has more advantages in controlling the overall execution cost of workflow tasks. Because of its randomness, FIFO has no corresponding optimization in the execution cost, so its cost-effectiveness ratio is very low.

## 5. Conclusion

In order to solve the problem that the stability of the model will decline due to the interference of human factors in the process of multitask scheduling in the traditional algorithm, a reasonable scheduling method based on the priority principle in company management based on the multi-constraint scheduling model is proposed. This paper expounds the principle of multitask scheduling in company management, constructs a priority-based multitask scheduling network model based on the above principles,

conducts virtual scheduling on the model, searches the optimal solution in the solution space of the scheduling problem, and obtains the reasonable multitask scheduling method in the company management. The experimental results show that the improved algorithm can improve the resource utilization rate and scheduling efficiency and avoid the defect that the stability of the scheduling model is reduced in the case of multiwork doping, and the satisfactory effect is achieved. The use environment of the model is more complex, and the role mapping and permission inheritance between multiple domains are taken into account, so that it can be applied to a wider range of fields. The various types of the cloud computing simulation platform are continued to be studied and strived to expand them to better simulate task scheduling.

### Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

### Conflicts of Interest

The author declares that there are no known competing financial interest or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgments

This work was supported by the Jilin Social Science Fund Project: research on the influence mechanism of Confucian culture on corporate governance (no. 2018B71).

### References

- [1] A. Noroozi and H. Mokhtari, "Scheduling of printed circuit board (PCB) assembly systems with heterogeneous processors using simulation-based intelligent optimization methods," *Neural Computing and Applications*, vol. 26, no. 4, pp. 857–873, 2015.
- [2] S. Jia, G. Yan, A. Shen, and J. Zheng, "Dynamic simulation analysis of a construction and demolition waste management model under penalty and subsidy mechanisms," *Journal of Cleaner Production*, vol. 147, pp. 531–545, 2017.
- [3] H. Zhang, J. Xie, J. Ge, J. Shi, and W. Lu, "Optimization model and online task interleaving scheduling algorithm for MIMO radar," *Computers & Industrial Engineering*, vol. 127, pp. 865–874, 2019.
- [4] L. He, Y. Ren, Z. X. Feng et al., "Numerical simulation and design optimization Research of slope ratio and platforms influence on a high rock cut slope engineering stability," *Advanced Materials Research*, vol. 1065-1069, pp. 151–158, 2015.
- [5] O. Z. Sharaf and M. F. Orhan, "Thermodynamic analysis and optimization of densely-packed receiver assembly components in high-concentration CPVT solar collectors," *Energy Conversion and Management*, vol. 121, pp. 113–144, 2016.
- [6] Y. L. Xie, D. X. Xia, L. Ji et al., "An inexact stochastic-fuzzy optimization model for agricultural water allocation and land resources utilization management under considering effective rainfall," *Ecological Indicators*, vol. 92, pp. 301–311, 2017.
- [7] A. Veremyev, V. Boginski, and E. L. Pasiliao, "Potential energy principles in networked systems and their connections to optimization problems on graphs," *Optimization Letters*, vol. 9, no. 3, pp. 585–600, 2015.
- [8] M. K. Mehlawat, P. Gupta, and W. Pedrycz, "A new possibilistic optimization model for multiple criteria assignment problem," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 4, pp. 1775–1788, 2018.
- [9] B. Ai, C. C. Wang, and B. Y. Xu, "Optimization of initialization conditions in Aspen simulation process for batch distillation," *Modern Chemical Industry*, vol. 37, no. 12, pp. 190–192, 2017.
- [10] H. Vanharanta and E. Markopoulos, "The applied philosophy concept for management and leadership objects through the company democracy model," *Theoretical Issues in Ergonomics Science*, vol. 20, no. 2, pp. 178–195, 2019.
- [11] F. Zhang and J. Li, "An improved particle swarm optimization algorithm for integrated scheduling model in AGV-served manufacturing systems," *Journal of Advanced Manufacturing Systems*, vol. 17, no. 3, pp. 375–390, 2018.
- [12] M. Dobroselskyi and R. Madleňák, "Model of waste transportation management in the conditions of a production company," *Transportation Research Procedia*, vol. 40, pp. 1023–1029, 2019.
- [13] T. Ning and X. P. Wang, "Study on disruption management strategy of job-shop scheduling problem based on prospect theory," *Journal of Cleaner Production*, vol. 194, pp. 174–178, 2018.
- [14] C. C. David, L. C. D. José, J. M. R. Javier et al., "Self-management model in the scheduling of successive appointments in rheumatology," *Reumatologia Clinica*, vol. 15, no. 6, pp. 333–337, 2018.
- [15] H. Cornelius, S. Zhichao, R. M. Oliveira et al., "Knowledge-aided informed dynamic scheduling for LDPC decoding of short blocks," *Iet Communications*, vol. 12, no. 9, pp. 1094–1101, 2018.
- [16] S. Carrye, "Invisible disabilities: perceptions and barriers to reasonable accommodations in the workplace," *Library Management*, vol. 40, no. 1-2, pp. 113–120, 2019.
- [17] B. Jiang, C. J. Tang, and C. Yan, "A stochastic programming model for outpatient appointment scheduling considering unpunctuality," *Omega*, vol. 82, pp. 70–82, 2019.
- [18] M. Chen, S. Lu, and Q. Liu, "Uniform regularity for a Keller-Segel-Navier-Stokes system," *Applied Mathematics Letters*, vol. 107, Article ID 106476, 2020.
- [19] Y. Qin, G. Zeng, R. Kurachi et al., "Energy-efficient intra-task DVFS scheduling using linear programming formulation," *IEEE Access*, vol. 7, pp. 30536–30547, 2019.
- [20] C. Zeng, J. Tang, and Z. P. Fan, "Auction-based cooperation mechanism for cell part scheduling with transportation capacity constraint," *International Journal of Production Research*, vol. 57, no. 11-12, pp. 3831–3846, 2019.
- [21] K. Cui and X. Jing, "Research on prediction model of geotechnical parameters based on BP neural network," *Neural Computing and Applications*, vol. 31, no. 12, pp. 8205–8215, 2019.
- [22] S. Rubab, M. F. Hassan, A. Mahmood et al., "QoS based multi-constraints bin packing job scheduling heuristic for heterogeneous volunteer grid resources," *The International Arab Journal of Information Technology*, vol. 16, no. 4, pp. 661–668, 2019.
- [23] A. Syed and G. Fohler, "Efficient offline scheduling of task-sets with complex constraints on large distributed time-triggered systems," *Real-Time Systems*, vol. 55, no. 2, pp. 209–247, 2019.



## Research Article

# A 3D Reconstruction Method Using Multisensor Fusion in Large-Scale Indoor Scenes

Panlong Gu <sup>1</sup>, Fengyu Zhou <sup>1</sup>, Dianguo Yu <sup>2</sup>, Fang Wan <sup>1</sup>, Wei Wang <sup>3</sup>,  
and Bangguo Yu <sup>1</sup>

<sup>1</sup>School of Control Science and Engineering, Shandong University, Jinan250061, Shandong, China

<sup>2</sup>Institute of Technology, Qu Fu Normal University, Jining 276825, Shandong, China

<sup>3</sup>Faculty of Computer Science and Technology, Qi Lu University of Technology, Jining 276825, Shandong, China

Correspondence should be addressed to Fengyu Zhou; [zhoufengyu@sdu.edu.cn](mailto:zhoufengyu@sdu.edu.cn)

Received 23 April 2020; Revised 11 August 2020; Accepted 9 September 2020; Published 24 September 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Panlong Gu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

RGBD camera-based VSLAM (Visual Simultaneous Localization and Mapping) algorithm is usually applied to assist robots with real-time mapping. However, due to the limited measuring principle, accuracy, and distance of the equipped camera, this algorithm has typical disadvantages in the large and dynamic scenes with complex lightings, such as poor mapping accuracy, easy loss of robot position, and much cost on computing resources. Regarding these issues, this paper proposes a new method of 3D interior construction, which combines laser radar and an RGBD camera. Meanwhile, it is developed based on the Cartographer laser SLAM algorithm. The proposed method mainly takes two steps. The first step is to do the 3D reconstruction using the Cartographer algorithm and RGBD camera. It firstly applies the Cartographer algorithm to calculate the pose of the RGBD camera and to generate a submap. Then, a real-time 3D point cloud generated by using the RGBD camera is inserted into the submap, and the real-time interior construction is finished. The second step is to improve Cartographer loop-closure quality by the visual loop-closure for the sake of correcting the generated map. Compared with traditional methods in large-scale indoor scenes, the proposed algorithm in this paper shows higher precision, faster speed, and stronger robustness in such contexts, especially with complex light and dynamic objects, respectively.

## 1. Introduction

In order to build intelligent applications for mobile robots in a large scene, autonomous and efficient navigation plays an increasingly important role. VSLAM, a technique with high-precision positioning and obstacle avoidance information, has been popular in the field of robot navigation. At present, VSLAM is usually applied on monocular camera, binocular camera, and depth camera RGBD. Among these three, the RGBD-based VSLAM algorithm has obvious advantages in terms of real-time accuracy. This is mainly because of the fact that the depth camera can input the depth image with accurate scale information to the computer directly. It not only overcomes the problem that the binocular camera needs to consume extra computer resources to obtain the depth image, but also solves the problem that the monocular

camera cannot do dense construction during mapping, or has no scale information and poor drawing quality in mapping. Despite of advantages, some popular RGBD-based VSLAM algorithms, such as RGBD SLAM V2 algorithm [1], RTAB algorithm [2, 3], and DVO algorithm [4], still have typical limitations in large-scale indoor scenes, which mainly reflect in that dynamic objects will introduce additional errors to VSLAM or even cause to operation failure. However, this would be impossible in real-time large scene reconstruction. For example, some inevitable changes, lights, people moving, and object moving can cause the loss of position and posture of mobile robots [5] and, in turn, lead to inaccurate mapping. In order to deal with some dynamic objects that may exist in large scenes, the current algorithm usually eliminates dynamic objects by adding object detection or image segmentation algorithm based on

deep learning in the system, such as Dynamic SLAM [6], Cluster VO [7], and Cluster SLAM [8] algorithm [9]. However, this will inevitably lead to a large consumption of computing resources, and some microdynamic objects are usually distributed near the camera. After removing these feature points, the accuracy of the camera odometer will also be affected. In addition, since the effective measuring distance of the RGBD depth camera (i.e., Kinect V2) is only 4.5 meters [5], it may result in a large cumulative error in the large scenes. Therefore, it is challenging to obtain a high-precision environmental map. Nevertheless, this weakness can be compensated by using a 2D laser range finder. For example, the commonly used SICK TIM 561 laser range finder has 270 degree field of view angle with 12 meter effective measuring distance. Cartographer [10], a laser SLAM algorithm based on graph optimization proposed by Google in 2016, is specifically used on such kind of a 2D laser range finder. It applies SPA [10, 11] (Sparse Pose Adjustment), as well as the branch and bound method [10, 12], to adjust the constructed map and reduces the accumulated error effectively. Hence, it is more suitable for large-scale scene mapping. However, this method still has some shortcomings. A typical weakness lies in its limitation of data that is obtained only from a certain level of the space. This limitation may lead to the loss of large amount of environmental information during the performance of mobile robots and can hamper their behaviors and control in turn.

According to the analysis mentioned above, based on the existing Cartographer algorithm, this paper proposes a new method for 3D mapping in large scene contexts, which integrates a laser range finder and RGBD depth camera. Its working principles mainly include three steps. Firstly, it inserts the 3D point cloud generated in real-time into the submap generated by the Cartographer algorithm for 3D reconstruction. Then, visual loopback is introduced to check the accuracy of laser loopback. It needs to run a visual loopback test on the back end of the computer (not in real-time), which will check the loopback results established by the Cartographer algorithm. If the two loopback results are significantly different, it will replace the laser loopback with the results obtained by the visual loopback. Finally, after the algorithm detects the loop-closure, it adjusts the pose of the point cloud in real-time, modifies the map, and inserts the 3D point cloud bound to the submap into the appropriate space position. Comparative experiment results show that the proposed algorithm is more accurate, real-time, and robust than the simple vision SLAM and laser SLAM algorithms.

The main contributions of this paper can be summarized into three aspects:

- (i) The approach proposed in this paper about reconstruction in large dynamic scene implements the fusion of RGBD camera and laser range finder based on the Cartographer algorithm. Cartographer provides more robust odometer data for algorithm reconstruction, and an RGBD camera provides rich environment data for the algorithm to eliminate the point clouds and laser points of dynamic objects, so

as to optimize the odometer calculation of Cartographer.

- (ii) A novel approach to calculate the ground true value in large scene, which does not need extra instruments but simply laying the Apriltag on the ground instead. With the time and cost required in the experiment being greatly reduced.
- (iii) Experimental results show that the proposed approach has achieved the robust state-of-the-art performance on the real large public scene.

The remainder of this paper is organized as follows: Section 2 presents the related work in the past few years about 3D reconstruction in dynamic scene and the Cartographer algorithm presented in this paper. Section 3 describes the sensor calibration method in our proposed approach. Section 4 introduces the new approach in detail. The experimental results in the real scene are given in Section 5. Conclusions are drawn in Section 6.

## 2. Related Work

This section focuses on the review of previous works in terms of the 3D reconstruction method of dynamic scale and the introduction of Cartographer algorithm, upon which our new method is developed.

*2.1. 3D Reconstruction in Large Dynamic Scenes.* To avoid generating errors while reconstructing, existing works addressing this problem are of three types.

The first category is based on dynamic object detection. In [9, 13], algorithms reduce the introduced error by removing the feature points in the image where the dynamic object is located and, then, uses the points in the remaining static objects to calculate the camera odometer data of the two frames.

The second category is based on the random consistency of established maps. In [14], the random consistency algorithm was proposed to evaluate the established map, which, in turn, removes the dynamic objects mistakenly inserted into the map.

The third category is based online and surface features. References [15, 16] proposed methods to calculate the pose change of the camera between two frames by line and surface features instead of point features.

Although these algorithms can solve the interference of moving objects in the scene, they will increase extra burden on computing and bring about disadvantageous effects on the accuracy and the real-time performance. Especially when the proportion of external interference data exceeds a certain threshold [17], no algorithm to remove noise or interference could help.

*2.2. Cartographer Algorithm.* The Cartographer algorithm [10], a cross-platform and multisensor fusion -based laser

SLAM algorithm proposed by Google, consists of local SLAM and global SLAM algorithms.

In the local SLAM algorithm, Cartographer calculates odometer information by fitting scan points and IMU data from newly inserted scan points. Calculating the position of the newly inserted scan point in the submap is the least square problem [10] which aims to maximize the probability of matching the scan point with the submap.

The submap generated by local SLAM is a part of the global map, which is presented in the form of probability grid. Each grid has a fixed odds value, which indicates the blocked probability of the grid. After the new scan points are inserted into the submap, a set of odds are recalculated, and the probability values  $p_{\text{hits}}$  or  $p_{\text{miss}}$  of the grid points represented by odds set is updated using probability formula (1), Reference [10], and the score of each point of the probability grid is calculated by formula (2). After sufficient scans have been inserted, the current submap will be updated and output.

$$\text{odd}(p) = \frac{p}{1-p}, \quad (1)$$

$$M_{\text{new}}(x) = \text{clamp}(\text{odd}^{-1}(\text{odds}(M_{\text{old}}(x)) \cdot \text{odds}(p_{\text{hits/misses}}))). \quad (2)$$

In the global SLAM algorithm, Cartographer recorded each submap generated by the local SLAM algorithm and the scans data and their locations for loop-closure detection. During the movement of the robot, the branch and bound method [12] was used to establish loop-closure constraints, using SPA (Sparse Pose Adjustment) algorithm [11] to optimize the locations of all submaps.

With the assistance of global SLAM and local SLAM, the Cartographer algorithm has much better robustness, real-time performance, and accuracy in the real scene than the compared methods.

### 3. Sensor Calibration

In this paper, the large scene reconstruction algorithm integrates an RGBD depth camera and laser range finder. Due to the need to superimpose the point cloud data of the RGBD camera according to the laser range finder, odometer sensor calibration is the first and the necessary step to do. The proposed algorithm has two key components: (i) point cloud generation and (ii) registration of point cloud to the laser range finder coordinate system.

**3.1. Point Cloud Generation.** The RGBD camera to be used in this paper is Kinect  $v_2$ , and its depth camera and RGB camera are located in different positions of Kinect  $v_2$ . In order to obtain accurate environmental point cloud, it is necessary to calibrate the depth camera first.

Firstly, we can get the RGB camera's internal parameter matrix  $C_c$ , depth camera's internal parameter matrix  $C_d$ , and the external parameter matrix  $M$  of these two cameras by Zhang's calibration method [18], respectively.

Then, we assume that the points in the space are at the same distance from the center of the two cameras and substitute matrix  $C_c$ ,  $C_d$ , and  $M$  in the derived equation as follows:

$$\begin{bmatrix} Dx_c \\ Dy_c \\ 1 \\ \frac{1}{d} \end{bmatrix} = C_c \cdot M \cdot C_d^{-1} \begin{bmatrix} Dx_d \\ Dy_d \\ 1 \\ \frac{1}{d} \end{bmatrix}, \quad (3)$$

where  $d$  is the pixel value at the depth coordinate  $(Dx_d, Dy_d)$  and denotes the depth of this point. By equation (3), point  $(Dx_d, Dy_d)$  on the depth image can be transformed to the RGB image coordinate system and recorded as  $(Dx_c, Dy_c)$ .

Finally, this method is used to transform each point on the depth image to the RGB image coordinate. By superposing to display the depth image with the color image, as can be seen in Figure 1, the color image and the depth image are well coincided.

After calibration, the depth  $d$  of each point  $P_{\text{img}} = (x_{\text{img}}, y_{\text{img}})$  of the RGB image can be obtained by using the corresponding pixel value in the calibrated depth image transformed, then the coordinate  $P_{\text{img}}$  and depth  $d$  are plugged into equations (4)–(6) [19, 20], and the position of this point  $P_{\text{cam}} = [P_x, P_y, P_z]$  is figured out in space.

$$P_z = d, \quad (4)$$

$$P_x = \frac{P_z}{f_x} (x_{\text{img}} - c_x), \quad (5)$$

$$P_y = \frac{P_z}{f_y} (y_{\text{img}} - c_y). \quad (6)$$

In the abovementioned equations,  $(f_x, f_y, c_x, c_y)$  expresses the parameter in the camera's internal parameter matrix focal length of the depth camera. The parameter represents the optical center coordinate and the product of the camera zoom factor and focal length, respectively.

**3.2. Registration of Point Cloud to the Laser Range Finder Coordinate System.** After obtaining the point cloud, in order to accurately insert the point cloud data into the grid map generated by Cartographer, the camera point cloud and laser range finder scanning points need to be calibrated. The position of the calibration plate relative to the robot is shown in Figure 2.

The point cloud in world coordinate  $P_{\text{world}}$  and its corresponding points in the RGBD camera coordinate  $P_{\text{cam}}$  system can be expressed by the following equation:

$$P_{\text{cam}} = R P_{\text{world}} + t. \quad (7)$$

The  $3 \times 3$  rotation matrix  $R$  and 3-vector translation  $t$  represent the orientation and the position of the camera,

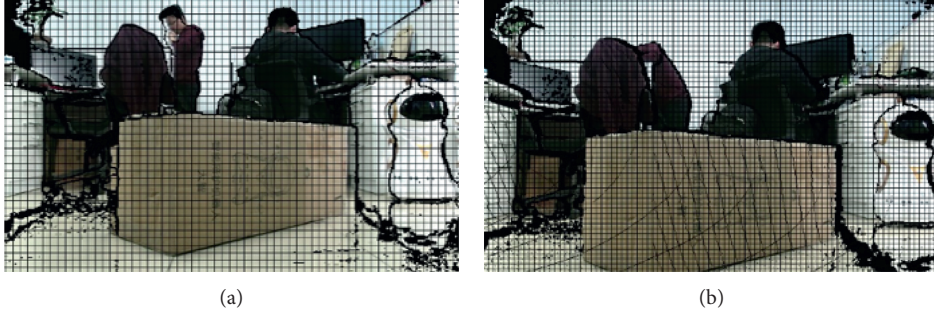


FIGURE 1: Calibration results between RGB and depth of Kinect. (a) Camera data cannot be aligned before calibration. (b) After calibration, the depth image can be well-aligned with the color image.



FIGURE 2: Calibration is performed using a calibration plate affixed to a prominent raised surface.

respectively. Similarly, it can be deduced that each scanning point  $P_{\text{laser}}$  measured by the laser range finder in the laser coordinate system has a corresponding point  $P_{\text{cam}}$  in the camera coordinate system, and the transformation relationship between them can be expressed by the following equation:

$$P_{\text{laser}} = \Phi P_{\text{cam}} + \Delta. \quad (8)$$

In equation (8),  $\Phi$  is a  $3 \times 3$  orthogonal matrix, representing the rotation relationship between the reference system of laser range finder and the world reference system.  $\Delta$  is a translation 3-vector representing the relative displacement between the two reference systems.

In order to calibrate the laser range finder and RGBD camera, there are two steps to do. Firstly, assuming the calibration board is located in the plane of  $Z=0$  in the world coordinate system, the surface of the calibration plate identified in the camera image is transformed into a 3-vector matrix  $N$  [20]. The norm of the matrix is adopted to show the Euclidean distance from the camera phase center to the calibration board. According to the transformation relationship between the camera coordinate system and real coordinate system as expressed in equation (7), the vector matrix can be deduced by the following equation:

$$N = -R_3(R_3^T \times t). \quad (9)$$

Secondly, assuming that the plane detected by the 2D laser range finder is located at  $Y=0$  in the world coordinate, the scanning points recorded by the laser rangefinder on the chessboard are represented by the following equation:

$$\hat{P}_{\text{laser}} = [X, Z, 1]^T. \quad (10)$$

Substituted  $\hat{P}_s$  into equation (8), the corresponding points of the scanning point in the camera reference can be obtained. When  $\hat{P}_{\text{laser}} = [X, Z, 1]^T$  is located in the calibration plane represented by the parameter  $N$ ,

$$\|N\|^2 = N \cdot P. \quad (11)$$

Furthermore, equation (11) can be deduced as equation (12):

$$\|N\|^2 = N \cdot H\hat{P}_s, \quad (12)$$

$$H = \Phi^{-1} \begin{bmatrix} 1 & 0 & -\Delta_1 \\ 0 & 0 & -\Delta_2 \\ 0 & 1 & -\Delta_3 \end{bmatrix}, \quad (13)$$

$$\Delta = [\Delta_1, \Delta_2, \Delta_3]^T. \quad (14)$$

After solving linear equation (12), the external parameter matrix from the RGBD camera to the laser range finder can be calculated. Then, we substitute  $\Phi$  and  $\Delta$  into equation (8) and transfer each laser scanning point. The calibration between the  $s$  scanning point and the point cloud can be calibrated. The calibration effect is shown in Figure 3.

## 4. System Overview

Figure 4 is the overall architecture of the proposed algorithm in this paper, which is divided into two major parts: the Cartographer algorithm and 3D reconstruction.

The Cartographer algorithm consists of local SLAM and global SLAM algorithms and provides odometer information and submap information for 3D reconstruction. The local SLAM is used to fit the acceleration and angle data of the newly inserted laser scanning point and IMU (inertial measurement unit) compare with the laser scanning points of the previous frame in order to calculate the acceleration, angle data, and odometer information. Also, the pose of the generated submap representing the new area is continuously adjusted, to which the robot reaches. The global SLAM algorithm, or global optimization algorithm, applies the branch and bound method to perform closed-loop matching on the newly generated submap with all previous submaps and point set data.

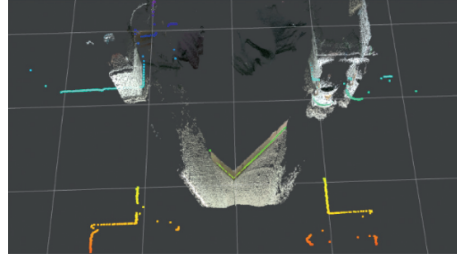


FIGURE 3: Calibration results between Kinect and the laser range finder.

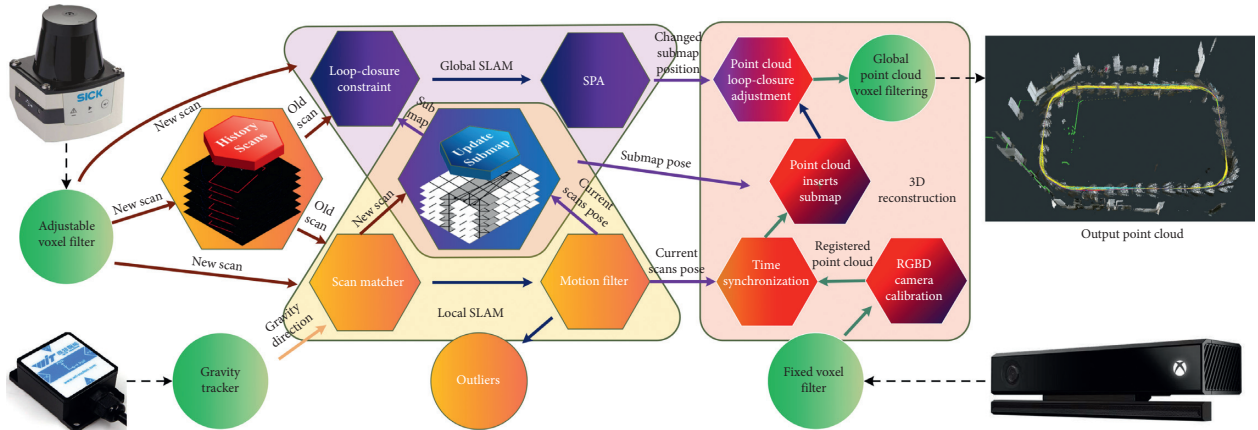


FIGURE 4: Overview of the 3d scenario construction system.

It uses SPA (Sparse Pose Adjustment) to adjust all previously generated submaps and scan point sets. After the match completes, the construction of 2D map is performed.

In the part of 3D reconstruction, after receiving the data of the depth camera, firstly, the corresponding point cloud information is calculated, and the position and pose are transformed and bound to the corresponding submap space. Along with the adjustment of the submap pose by Cartographer, the point cloud data bound to the submap is adjusted in real-time simultaneously. When superposition submap generates a 2D map, superposition point cloud generates a 3D map and publishes point cloud information to the robot in real-time.

In more detail, the proposed algorithm is introduced from three major functional parts which are described as follows.

**4.1. Filling Space Point Cloud into the Submap.** The proposed algorithm builds the map by filling the space point cloud into the submap. In order to accurately superimpose the 3D point cloud information that is inserted into each submap, the point cloud information needs to be bound with each submap. Then, a method with less resource consumption needs to be found to update the adjusted point cloud. Along with robot moving forward and continuous insertion of laser scanning points, the Cartographer algorithm will estimate the robot's mobile posture among scanning points and output the robot pose at the frequency of 70 Hz as shown in the following equation:

$$\delta = [\delta_x, \delta_y, \delta_z, \delta_{roll}, \delta_{pitch}, \delta_{yaw}]^T. \quad (15)$$

When the robot reaches a new area, the algorithm will generate a submap and publish all the previously established submaps' information at the frequency of 5 Hz, and the pose of submap can be expressed by the following equation:

$$\phi = [\phi_x, \phi_y, \phi_z, \phi_{roll}, \phi_{pitch}, \phi_{yaw}]^T. \quad (16)$$

At the same time, the Kinect depth camera on the robot publishes depth images and RGB images at a frequency of 25 Hz, and the IMU publishes robot pose information at a frequency of 200 Hz.

In ROS, computer uses time stamp to record time information when data are published. The time stamps of the robot pose, submap, RGBD camera image, and IMU are represented by  $T_\delta$ ,  $T_\phi$ ,  $T_{RGBD}$ , and  $T_{IMU}$ , respectively.

In order to ensure the normal operation of the algorithm, different sensors need to be synchronized to ensure the sensor data input to the system is of synchronization. Time synchronization includes hardware synchronization and software synchronization. Hardware synchronization is achieved using the BIOS time to screen multiple sensors' data, but cannot be achieved among different devices. Software synchronization caters for this need. This type of synchronization uses software to select data closest to the base time from the sensor data that are input to the same host and completes the time binding of sensor data. Main steps are taken as follows:

The first step is to synchronize the pose of the RGBD camera and the pose of robot, which is to find the pose data closest to  $T_{\text{RGBD}}$  in the prestored 2000 pose data of the robot, as the process shown in Algorithm 1.

Similarly, such synchronization can be achieved for the RGBD camera and the submap, by which the point cloud data closest to  $T_{\phi}$  can be selected.

In the process of binding RGBD image data, robot pose information, and IMU data, the maximum theoretical time difference is 7 ms while the average time difference in actual tests is less than 1 ms. If the robot runs at the speed of 2 m/s, the distance error generated is of the millimeter level. In the process of synchronizing submap and RGBD point cloud, the time difference in theory is 20 ms if the release frequency of the RGBD image is at 25 Hz. In real running, the measured time difference is about 100 ms due to the time consumption on point cloud generation. When the robot is running at 2 m/s, the center of the submap and the point cloud at a distance deviation of 20 cm will be generated if the point cloud is directly inserted into the submap. The abovementioned error data generated in time synchronization are shown in Table 1.

In order to eliminate the abovementioned spatial dislocation, it is necessary to transfer the point cloud to the reference frame of its corresponding submap before the point cloud is filled into the submap space. To achieve this, it is, firstly, needed to calculate the spatial position difference between the robot pose and the submap pose while generating the point cloud and, then, make use of equation (22) to transform the change of the rotation angle between two reference systems into rotation matrix  $R$ [21], which can be calculated by equations (17)–(21).

$$\tau = [\delta_x - \phi_x, \delta_y - \phi_y, \delta_z - \phi_z], \quad (17)$$

$$\gamma = [\delta_{\text{roll}} - \phi_{\text{roll}}, \delta_{\text{pitch}} - \phi_{\text{pitch}}, \delta_{\text{yaw}} - \phi_{\text{yaw}}], \quad (18)$$

$$R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \gamma_{\text{roll}} & -\sin \gamma_{\text{roll}} \\ 0 & \sin \gamma_{\text{roll}} & \cos \gamma_{\text{roll}} \end{bmatrix}, \quad (19)$$

$$R_y = \begin{bmatrix} \cos \gamma_{\text{pitch}} & 0 & \sin \gamma_{\text{pitch}} \\ 0 & 1 & 0 \\ -\sin \gamma_{\text{pitch}} & 0 & \cos \gamma_{\text{pitch}} \end{bmatrix}, \quad (20)$$

$$R_z = \begin{bmatrix} \cos \gamma_{\text{yaw}} & -\sin \gamma_{\text{yaw}} & 0 \\ \sin \gamma_{\text{yaw}} & \cos \gamma_{\text{yaw}} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (21)$$

$$R = R_z R_y R_x, \quad (22)$$

$$\text{Trans} = \begin{bmatrix} R & \tau \\ 0^T & 1 \end{bmatrix}. \quad (23)$$

After obtaining the rotation matrix and translation matrix,  $R$  and  $\tau$  will be taken into equation (23). According

```

Procedure Time synchronized
 $T_{\text{base}} \leftarrow T_{\text{RGBD}}$ 
 $\text{data}[2000] \leftarrow \text{Sensor\_data}(\text{aim\_sensor})$ 
 $\text{mini\_time} \leftarrow \text{max\_timeinterval}()$ 
for  $i$  in 2000, do
   $T_{\text{aim}} \leftarrow \text{data}[i]$ 
  if  $\text{abs}(-T_{\text{aim}}) < \text{threshold}$ , then
     $\text{minitime} \leftarrow \text{abs}(-)$ 
     $\text{mini\_I} \leftarrow i$ 
  end if
end for
 $\text{aimsensordata} \leftarrow \text{data}[i]$ 
return  $\text{aimsensordata}$ 

```

ALGORITHM 1: Time synchronized.

TABLE 1: Errors occurred in time synchronization.

| Objects that synchronize with RGBD | IMU  | Submap |
|------------------------------------|------|--------|
| Theoretical time error (ms)        | 7    | 20     |
| Theoretical distance error (cm)    | 1.4  | 4      |
| Actual time error (ms)             | <1   | >100   |
| Actual distance error (cm)         | <0.2 | >20    |

to equation (24), all points in the original point cloud, denoted by  $P$ , can be inserted into the submap, denoted by  $P_{\text{submap}}$ .

$$P_{\text{submap}} = P \cdot \text{Trans}. \quad (24)$$

**4.2. Point Cloud Superposition.** The pose of the submap needs to be adjusted continuously in the process of loop optimization. To facilitate the adjustment, on top of the transformation mentioned in part A, it still needs to transfer the pose of the point cloud to its corresponding spatial position. This requires a translation of the point cloud using equation (25) based on the submap's location information.

$$\tau_{\text{submap}} = [\phi_x \ \phi_y \ \phi_z]^T. \quad (25)$$

After that, the translated point cloud will be superimposed on the total point cloud. The Cartographer algorithm constantly generates submaps when running.

With the superposition of submaps, the bound point clouds are also superposed. In the process of generating the 2D map, the joint of 3D point cloud can be completed. Figure 5 gives an example. Along with the robot moving forward, the Cartographer algorithm generates submap14 and submap15. When the two submaps are superposed, the corresponding point clouds of these two submaps are also superposed.

**4.3. Dynamic Object Removal.** In the process of 3D reconstruction, there are usually a large number of dynamic objects in the scene, and it will not only affect the accuracy of odometer but also introduce dynamic noise points into map.

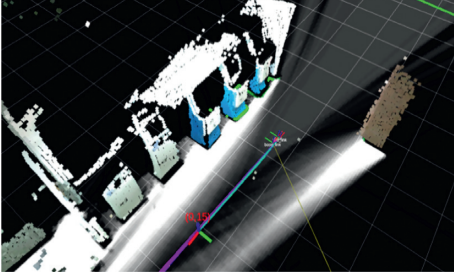


FIGURE 5: Binding of point cloud data to the submap.

The proposed algorithm based on the odometer calculated by the Cartographer algorithm has higher accuracy than VSLAM, in the dynamic scene.

However, the more accurate odometer can ensure the accuracy of the map, but it cannot avoid the residual dynamic point cloud being inserted into the map in the process of reconstruction. Therefore, in order to solve this problem, this section proposes a method to eliminate the residual dynamic point cloud with the least computing resources in the reconstruction.

In this algorithm, we calculate the relative pose of two submaps and their corresponding point clouds relative to submap (16) and (17). Then, the view intersection area of two cloud points is calculated based on this pose.

When the algorithm inserts a new point cloud into the total point cloud using formula (26), it first opens a window around each point of the new point cloud, runs, “AND” operates with the points in the total point cloud so that the points that exist in both point clouds are preserved.

In this way, the residual point cloud in the map can be eliminated, as shown in Figure 6.

**4.4. Adjustment of Point Cloud Sloop-Closure.** Along with the continuous insertion of new scanning points, the global SLAM algorithm will open a window around the scanning points for detection with the help of the branch and bound method.

However, the laser range finder can only obtains the data of a certain plane in space, which is limited for the Cartographer algorithm to calculate loop-closure precisely or establish a loop-closure in the position where laser features are scarce.

We designed to use the ORB feature points to extract point cloud feature in every submap. When a new submap is generated by local SLAM, our algorithm will detect the loop-closure between the current submap’s point cloud and the previous submap’ point cloud. When the visual loop and the laser loop are established at the same time, we use the loop information calculated by the visual loop to replace the laser loop information. When visual loop-closure was established without laser loop-closure established, visual loop-closure will force the Cartographer algorithm to establish loop-closure between submaps.

During the course of loop-closure detection, no matter whether the algorithm detects a loop-closure or not, SPA will be performed to adjust the pose of the submap, in which the

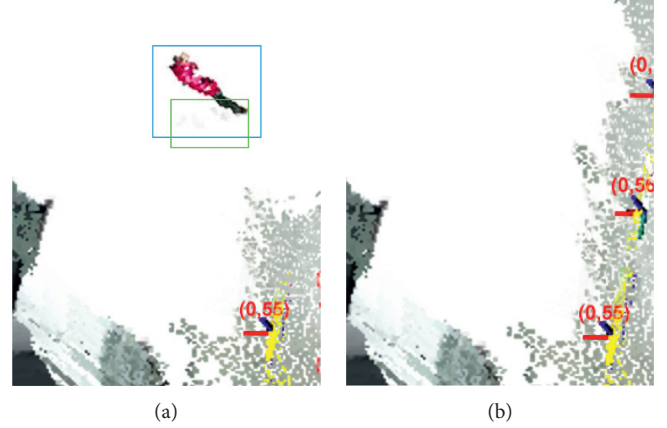


FIGURE 6: Dynamic object removal. (a) Before removal, the foot of the woman in red is detected, leaving a gray occupying grid and point cloud in the generated submap. (b) After removal, the woman in red walked through the machine and not appeared on the point cloud again. The Cartographer algorithm used formula (2) to update the occupied grid of basket position in (a) to the non-occupied state, and the proposed algorithm in section C removes the point cloud of the woman in red.

loop-closure occurs, from  $\phi_{old}$  to  $\phi_{new}$ . Through calculating the position difference of the old and the new submap, we can obtain the following equations:

$$\tau_{loop} = [\phi_{newx} - \phi_{oldx}, \phi_{newy} - \phi_{oldy}, \phi_{newz} - \phi_{oldz}], \quad (26)$$

$$\gamma_{loop} = [f_{newroll} - f_{oldroll}, \delta_{newpitch} - f_{oldpitch}, \delta_{newyaw} - f_{oldyaw}]. \quad (27)$$

Taking equations (26) and (27) into equations (22) and (23), the Euclidean transformation matrix between the new and the old submaps can be calculated, which is recorded as  $Trans_{loop}$ . Then, the loop-closure adjustment of the point cloud can be made by multiplying the  $Trans_{loop}$  and the point cloud bound by the submap that the pose has changed based on formula (24) and, then, translating and superposing the new generated point cloud according to formulas (25) and (26). Whenever a new submap is inserted, SPA will optimize and adjust the pose  $\phi$  of all generated submaps. Due to this reason, it is easy to jam the program and affect the display effect by resuperimposition of the point cloud. Moreover, after the superimposed point cloud is filtered by the voxel, the point cloud bound by the submap that the pose has been changed cannot be directly subtracted from the superimposed point cloud.

To solve the abovementioned problem, our algorithm divides the point cloud into several regions based on a certain threshold, and only the point clouds in the regions, where the submaps adjust their pose is located, are superimposed. The above algorithms are shown as (Algorithm 2)

## 5. Experiment

**5.1. Experiment Platform and Design.** In this paper, a Mecanum-wheel mobile robot equipped with a Kinect V2

```

Procedure Regional point cloud stack
Submap  $\leftarrow$  SubmapMessage
for  $i$  in size of (Submap), do
  if position(Submap[ $i$ ]) is changed and Submap[ $i$ ] is not new Submap then
    RegionalPointcloudUnfinish [IndexInRegional( $i$ )]
       $\leftarrow$  Equation (25) (RegionalPointcloudUnfinish [IndexInRegional( $i$ )], NewPosition(Submap[ $i$ ]))
  end if
  else if Submap[ $i$ ] is new Submap then
    NewPointcloud  $\leftarrow$  TimeSynchronized(Submap)
    TransformedPointCloud  $\leftarrow$  Equation (23) (NewPointcloud)
    ShiftedPointCloud  $\leftarrow$  Equation (25) (TransformedPointCloud, position(Submap[ $i$ ]))
    RegionalPointcloudUnfinish.pushback(ShiftedPointCloud)
    ResidualPointcloud  $\leftarrow$  ShiftedPointCloud + ResidualPointcloud
    if sizeof(RegionalPointcloudUnfinish)  $\geq$  threshold then
      for  $j$  in range(threshold)
        RegionAllPointcloud  $\leftarrow$  RegionalPointcloudUnfinish[ $j$ ] + RegionAllPointcloud
      end for
      RegionPointcloud.pushback(RegionAllPointcloud)
      RegionalPointcloudUnfinish.clear()
      ResidualPointcloud.clear()
      return RegionAllPointcloud
    else if
      return RegionAllPointcloud + ResidualPointcloud
    end if
  end if
end if

```

ALGORITHM 2: Regional point cloud stack.

depth camera, SICK TIM561 laser range finder, and IMU is chosen as the experiment platform, as shown in Figure 7, to perform various tests and comparative experiments in large indoor scenes.

In the experiment, we choose a relatively accurate range between 0.2 m and 5.5 m of Kinect depth data for calculation and down sample the point cloud using a 0.03 m grid-size voxel filter. After the calculation, the point cloud is sent to RVIZ for display. Also, we define the starting point of the mapping work as the origin of the world coordinate system, that is, the position of No. 0 submap.

In addition, in order to verify the robustness and accuracy of the algorithm based on its real-time performance, we select RTAB algorithm for comparison because of its better accuracy and robustness than other popular algorithms, such as DVO SLAM algorithm and RGB-D SLAM V2 algorithm.

**5.2. Comparison of Algorithm Robustness.** In order to verify the robustness of the proposed algorithm, we select the outer of the corridor hall (an area of 27 m  $\times$  16 m) on the 2nd floor of the main building at our campus as the experimental scene, as shown in Figure 8(a). In the course of mapping, the robot moves around the selected area at a speed of 2 m/s for one and a half laps.

As shown in Figure 8(b), due to a glass wall at the entrance of the hall, there is serious ambient light interference. Under this circumstance, the RGBD camera will lose most of its depth information, and such lost can exert a

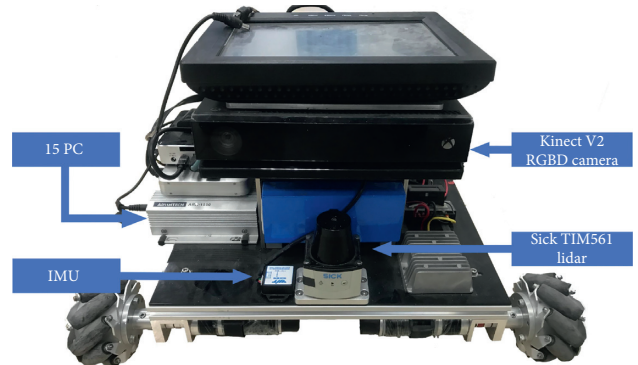


FIGURE 7: Experiment platform: the Meconium-wheel mobile robot.

great impact on RTAB algorithm's performance due to its sensitivity on visual information.

In many mapping experiments, RTAB algorithm is unable to complete the 3D mapping. The failures are all caused by the loss of location at the entrance. Despite that only a few mappings are successful, ghosting dislocation and distortion appear in many places, as shown in Figure 9(a).

Since RGBD's RTAB algorithm can lead to large error, our algorithm introduces a Cartographer algorithm, which reduces this error and improves the robustness, as shown in Figure 9(b).

**5.3. Comparison of Mapping Accuracy.** The lobby on the 1st floor of the selected building, which has sunlight interference





FIGURE 8: Second floor cloister of the main building. (a) Corridor hall on the 2nd floor. In the scene (b) with sunlight interference, depth camera (c) cannot obtain depth information.

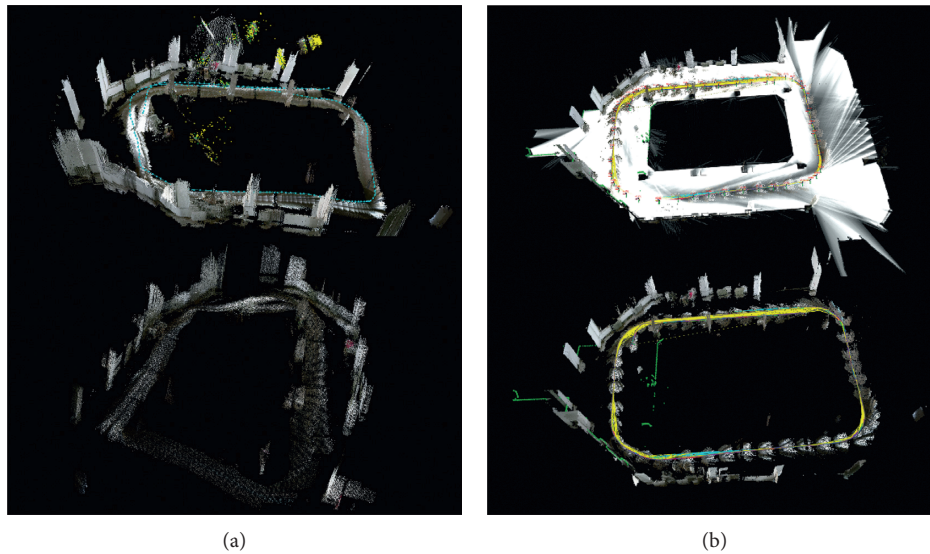


FIGURE 9: Comparison of mapping results on the second floor. (a) RTAB algorithm result. (b) Our result.

at night, sufficient indoor light source, and rich environmental characteristics, is selected as the experimental scene.

Its outer dimension is also  $27\text{ m} \times 16\text{ m}$ . The richer layout of the lobby, including several self-service recharge machines on both sides of the hall, many tables, and chairs, is the ideal place for comparative experiments, as shown in Figure 10. In addition, the scene has ideal conditions because there is no strong outdoor infrared interference.

The robot moves around the selected area of the first floor at the speed of  $2\text{ m/s}$  for one and a half laps, and the constructed 3D maps by our algorithm and RTAB algorithm are shown in Figures 11(a) and 11(b), respectively. It can be seen that both algorithms can complete the mapping in an ideal environment, without serious loss or damage to important environmental information.

Figures 11(c) and 11(d) are enlarged views of the 3D map constructed by our algorithm and RTAB algorithm, respectively. It can be seen that the map constructed by RTAB in Figure 11(d) has an obvious dislocation, while the mapping effect by our algorithm shown in Figure 11(c) is more accurate without significant dislocation.

A basic criterion of 3D mapping is no obvious error in point cloud superposition. For this reason, we measured

and compared the errors of the constructed 3D maps by both algorithms. Firstly, we use the laser range finder and meter ruler to measure the six specific locations in the field to obtain the actual size of the map, as shown in Figure 12.

In that, six numbers marked on the given map are used to represent the features of the selected six positions, which are (1) the length of self-service machine of campus card, (2) the horizontal distance between two bearing pillars of the main building hall, (3) the distance between the two walls at the entrance of the main building hall elevator, (4) the vertical distance between the main building hall's bearing columns, (5) the longest vertical distance of the main building hall, and (6) the longest horizontal distance of the main building hall.

Then, RVIZ is used to measure the length of the map constructed by two algorithms, and the absolute error and relative error of them are calculated, respectively, compared with the real value of the map. The comparison results are shown in Table 2.

As seen from Table 2, in terms of the relative errors, all relative errors of the maps constructed by our algorithm are less than 1%, which is much lower than that of the RTAB algorithm. In addition, the distribution of relative errors in



FIGURE 10: Environment of the lobby on the first floor. (a) The lobby on the 1st floor. (b) Self-service machine.

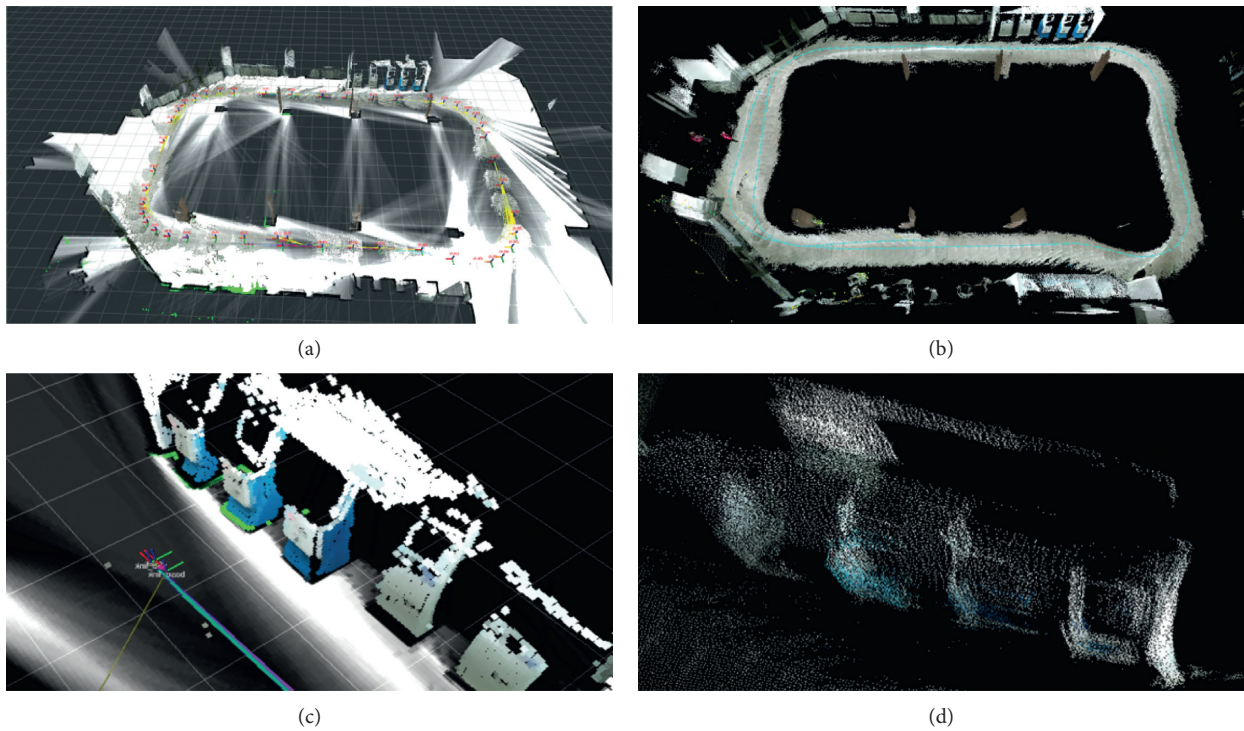


FIGURE 11: Comparison of the mapping effect on the first floor. (a), (b) The mapping effect of our proposed algorithm and RTAB, respectively. (c), (d) Enlarged details of the algorithm results.

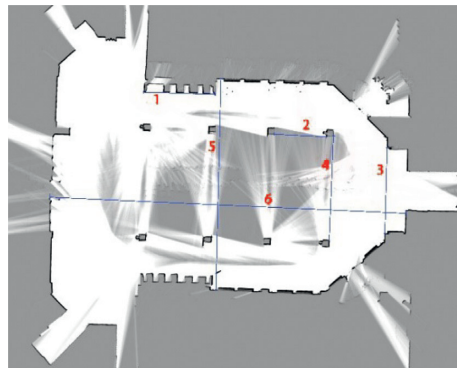


FIGURE 12: Mapping precision comparison.

TABLE 2: Measurement results of 3D mapping reconstruction.

| No. | Algorithm | GT (m) | Measurement (m) | Absolute error (m) | Relative error (%) |
|-----|-----------|--------|-----------------|--------------------|--------------------|
| 1   | OURS      | 5.51   | 5.53            | 0.02               | 0.363              |
|     | RTAB      |        | 5.41            | 0.10               | 1.815              |
| 2   | OURS      | 4.03   | 4.01            | 0.02               | 0.496              |
|     | RTAB      |        | 3.94            | 0.09               | 2.234              |
| 3   | OURS      | 6.52   | 6.50            | 0.02               | 0.307              |
|     | RTAB      |        | 6.36            | 0.16               | 2.454              |
| 4   | OURS      | 7.90   | 7.89            | 0.01               | 0.127              |
|     | RTAB      |        | 7.03            | 0.87               | 11.013             |
| 5   | OURS      | 16.05  | 15.98           | 0.07               | 0.436              |
|     | RTAB      |        | 15.56           | 0.49               | 3.053              |
| 6   | OURS      | 27.04  | 27.08           | 0.04               | 0.148              |
|     | RTAB      |        | 26.04           | 1.00               | 3.698              |

our algorithm is more even and is not affected by the sparse sensor feature points. However, the relative errors of RTAB algorithm are smaller in vertical scenes such as the position No. 3, No. 4, and No. 5 due to rich features, but larger in the case of No. 6 scene due to fewer feature points.

In terms of the absolute error, the results of our method are of relative balance, while the accuracy of RTAB algorithm increases along with the increase of map size and, in particular, reaches to 1 m in scene 6.

Based on the abovementioned comparisons, our algorithm is of significant advantages in a large scene.

#### 5.4. Comparison Experiment on Track and Loop Accuracy

##### 5.4.1. Calculating the Ground True Value of Robot Motion.

Experiments on the comparison of the track and the loop-closure precision require comparing the error between the ground truth value of robot motion and the errors of odometers in our method and the RTAB algorithm.

In the experiment, the ground truth value of robot motion is calculated by the position of Apriltag code on the floor tiles (60 cm × 60 cm per floor tile), as shown in Figure 13.

Firstly, the location of Apriltag code in the field is prestored in the industrial personal computer equipped by the robot using the following equation:

$$L_n = (x, y), \quad n = 1, 2, 3, \dots \quad (28)$$

Then, by identifying the Apriltag code laid on the ground, the position of the robot relative to Apriltag code is obtained, captured by the following equation:

$$L_A = (x, y). \quad (29)$$

Next, the position of the robot in the test site can be obtained by the following equation:

$$L_R = L_A + L_n. \quad (30)$$

According to the experiment of Wang and Olson [22], the robot's recognition accuracy of the Apriltag code can be up to 100% within 6 meters. Therefore, the ground true value



FIGURE 13: Apriltag on the ground.

of the robot's track can be calculated by the Apriltag algorithm.

##### 5.4.2. Contrast Experiments on Track and Loop-Closure Accuracy.

In order to compare the odometer accuracy and the loop-closure ability of two algorithms (our method and RTAB algorithm), we designed the following experiment. The remote-control robot first runs one lap along a rectangular track in the hall on the 1st floor, and continues running a certain distance after the robot passes the starting point in order to allow the algorithm to run loop-closure sufficiently. Finally, the robot returns to the starting point again.

Figures 14(a) and 14(b) show the comparison of the odometer. Here, the red asterisk and green asterisk in (a) represent the end point and the start point of the robot running, respectively, and (b) gives an enlarged figure showing that the proposed algorithm has much higher odometer accuracy in large-scale indoor scenes than that of RTAB algorithm.

To be more specific, as can be seen from Figure 14(b), the distance difference value between the start point and the end point of our method is 15.1 cm, while the data of RTAB is 65.307 cm.

Figures 14(c) and 14(d) show the comparison between the two algorithms after full loop-closure, and the result is that the loop-closure accuracy of our method is better than that of RTAB algorithm. As can be seen in (d) that shows more detailed information about the results, after loop-closure, most accumulated errors can be eliminated. In this case, the distance error in our algorithm is only 0.6 cm, while the result of RTAB algorithm is 1 cm.

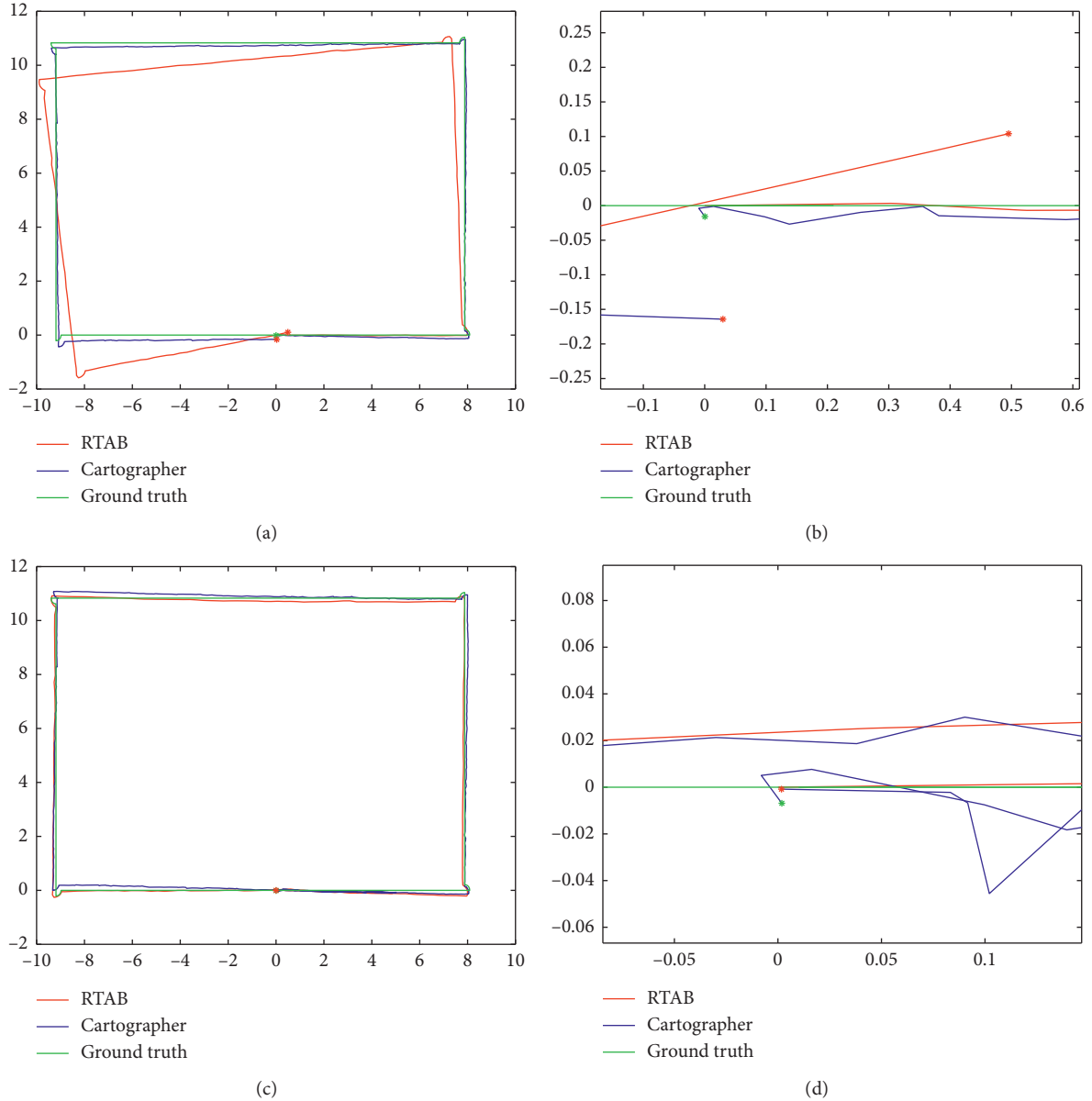


FIGURE 14: Trajectory accuracy contrast. (a) Accuracy comparison of the odometer. (b) Comparison of odometer accuracy after robot returns to origin. (c) Accuracy comparison of loop-closure. (d) Comparison of loop-closure accuracy after robot returns to origin. Unit: meter.

**5.5. Algorithm Contrast Experiment in Real-Time.** In order to ensure the integrity of the established map, in principle, our algorithm and RTAB algorithm are designed to process the current frame after all previously stored frames are processed. However, this design may not guarantee the real-time of map construction in reality.

In order to compare the real-time of two algorithms, in this paper, the environment information recorded by the robot in the real scene is packaged as rosbag function in ROS (robot operating system). During the verification phase, it is ensured that the sensor data received by the two algorithms are consistent through data packet playback in ROS.

Since the data types of the two algorithms are different, it is very troublesome and error-prone to record

and process each frame of the data produced by each algorithm. Instead, in the comparative experiment, we use the total time spent on mapping to compare the real-time performance of them. The experimental results are shown in Table 3.

It can be seen that the real-time of this algorithm is much higher than that of RTAB algorithm.

The computer configuration and environment used in this algorithm experiment are as follows:

CPU: I5-5200U 2.2GHZ

Ubuntu: 16.04

ROS: Kinetic

TABLE 3: The comparison of total time spent by the algorithms.

| Experiment site         | 1 <sup>st</sup> floor (s) | 2 <sup>nd</sup> floor (s) |
|-------------------------|---------------------------|---------------------------|
| Total rosbag duration   | 241                       | 213                       |
| Time for our algorithm  | 241                       | 213                       |
| Time for RTAB algorithm | 506                       | 421                       |

## 6. Conclusions

In the 3D reconstruction of a large scene, although the interference of moving objects and uneven sunlight can be eliminated, it needs to consume a huge amount of extra computing resources and, more importantly, is not conducive to the improvement of the real-time and accuracy of the algorithm. In this paper, a new algorithm is presented for 3D reconstruction in large scenes, which integrates the 3D point cloud generated by the Kinect camera and odometer information output by the Cartographer algorithm. Also, it is of high robustness and high accuracy. In addition, we also propose a new method applying Apriltag code to calculate the ground truth value of robot motion in large scenes. This method effectively solves the problem of the insufficient measurement range of the optical tracker in large scenes. The results of the real tests conducted in two selected actual scenes show that the newly proposed algorithm has higher real-time, robustness, and accuracy than the traditional RGBD camera-based 3D reconstruction algorithms.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported in part by the National Key R&D Program of China under Grant 2017YFB1302400, the National Natural Science Foundation of China under Grant nos. 61773242 and 61803227, and the Major Agricultural Applied Technological Innovation Projects of Shandong Province under Grant SD2019NJ014.

## References

- [1] F. Endres, J. Hess, J. Sturm, D. Cremers, and W. Burgard, "3-D mapping with an RGB-D camera," *IEEE Transactions on Robotics*, vol. 30, no. 1, pp. 177–187, 2014.
- [2] M. Labbé and F. Michaud, "RTAB-map as an open source lidar and visual SLAM library for large-scale and long-term online operation," *Journal of Field Robotics*, vol. 36, no. 2, pp. 416–446, 2019.
- [3] M. Labbé and F. Michaud, "Online global loop-closure detection for large-scale multi-session graph-based SLAM," in *Proceedings of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, Chicago, IL, USA, November 2014.
- [4] C. Kerl, J. Sturm, and D. Cremers, "Dense visual SLAM for RGB-D cameras," in *Proceedings of the 2013 IEEE/RSJ international conference on Intelligent Robots and Systems (IROS)*, January 2013.
- [5] K. Khoshelham and S. O. Elberink, "Accuracy and resolution of kinect depth data for indoor mapping applications," *Sensors*, vol. 12, no. 2, pp. 1437–1454, 2012.
- [6] L. Xiao, J. Wang, X. Qiu et al., "Dynamic-SLAM: semantic monocular visual localization and mapping based on deep learning in dynamic environment," *Robotics & Autonomous Systems*, vol. 17, pp. 1–16, 2019.
- [7] J. Huang, S. Yang, T. J. Mu, and S.-M. Hu, "ClusterVO: clustering moving instances and estimating visual odometry for self and surroundings," 2020, <https://arxiv.org/abs/2003.12980>.
- [8] J. Huang, S. Yang, Z. Zhao et al., "ClusterSLAM: a SLAM backend for simultaneous rigid body clustering and motion estimation," in *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, South Korea, February 2019.
- [9] L. Zhao, Z. Liu, J. Chen, W. Cai, W. Wang, and L. Zeng, "A compatible framework for RGB-D SLAM in dynamic scenes," *IEEE Access*, vol. 7, pp. 75604–75614, 2019.
- [10] W. Hess, D. Kohler, H. Rapp et al., "Real-time loop-closure in 2D LIDAR SLAM," in *Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA)*, June 2016.
- [11] K. Konolige, G. Grisetti, K. Rainer et al., "Efficient sparse pose adjustment for 2D mapping," in *Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, Taipei, Taiwan, October 2010.
- [12] J. Clausen, *Branch and Bound Algorithms-Principles and Examples*, pp. 1–30, Department of Computer Science, University of Copenhagen, Copenhagen, Denmark, 1999.
- [13] L. Zhang, L. Wei, P. Shen, W. Wei, G. Zhu, and J. Song, "Semantic SLAM based on object detection and improved octomap," *IEEE Access*, vol. 6, pp. 75545–75559, 2018.
- [14] C. Bibby and I. Reid, "Simultaneous localization and mapping in dynamic environments (SLAMIDE) with reversible data association," in *Proceedings of the Robotics: Science & Systems III, June, Georgia Institute of Technology*, Atlanta, GA, USA, 2007.
- [15] H. Li, Z. Hu, and X. Chen, "PLP-SLAM: a visual SLAM method based on point-line-plane feature fusion," *Robot*, vol. 39, no. 2, pp. 214–220+229, 2017.
- [16] Y. Liu, D. Yang, J. Li, Y. Gu, J. Pi, and X. Zhang, "Stereo visual-inertial SLAM with points and lines," *IEEE Access*, vol. 6, pp. 69381–69392, 2018.
- [17] D. Zou and P. Tan, "CoSLAM: collaborative visual SLAM in dynamic environments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 2, pp. 354–366, 2012.
- [18] Y. Liu and T. Li, "Reaserch of the improvement of Zhang's camera calibration method," *Optical Technique*, vol. 40, no. 6, pp. 565–570, 2014.
- [19] X. Jing, J.-L. Gou, X.-M. Ma, K. Huang, D. Liu, and Y.-M. Zhang, "A large viewing angle 3-dimensional V-SLAM algorithm with a kinect-based mobile robot system," *Robot*, vol. 36, no. 5, pp. 560–568, 2014.
- [20] Q. Zhang and R. Pless, "Extrinsic calibration of a camera and laser range finder (improves camera calibration)," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots & Systems*, Sendai, Japan, February 2005.

- [21] X. Gao and T. Zhang, *Visual SLAM 14: From Theory to Practice*, pp. 41–45, China Industry and Information Technology Publishing Group, Electronic Industry Press, Beijing, China, 2017.
- [22] J. Wang and E. Olson, “Apriltag 2: efficient and robust fiducial detection,” in *Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, October 2016.

## Research Article

# Meso-Complexity Computer Simulation Investigation on Antiexplosion Performance of Double-Layer Foam Aluminum under Pore Grading

Zhen Wang, Wen Bin Gu , Xing Bo Xie, Yu Tian Chen, and Lei Fu

Army Engineering University of PLA, Nanjing 210007, China

Correspondence should be addressed to Wen Bin Gu; 1120122090@bit.edu.cn

Received 12 July 2020; Revised 19 August 2020; Accepted 14 September 2020; Published 22 September 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Zhen Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Foam aluminum is an energy-absorbing material with excellent performance. The interlayer composed of multiple layers of foam aluminum and steel plate has good antiexplosion ability. In order to explore the antiexplosion performance of double-layer foam aluminum under different porosity rankings and to reveal its microscopic deformation law and failure mechanism, three kinds of aluminum foams with a porosity of 80%, 85%, and 90% were selected to form six different structures. Based on the Voronoi algorithm, a three-dimensional foam aluminum generation algorithm with random pore size and random wall thickness was written by using the Python language and Fortran language. The three-dimensional mesoscopic model of double-layer closed-cell aluminum foam sandwich panel is established by using LS-DYNA and ABAQUS software. The explosion process was simulated, and the flow field movement of explosion shock wave of aluminum foam under different porosity rankings was analyzed. Two groups of aluminum foam were randomly selected for the explosion test and compared for the strain and compression. The test results are consistent with the simulation results, which verifies the correctness of the three-dimensional meso-model. The results show that when the porosity of the upper layer of aluminum foam is greater than that of the lower layer of aluminum foam, the sandwich structure of double-layer aluminum foam has a large compression and the bottom plate has a small displacement; it is not that the greater the compression amount of aluminum foam is, the better the antiexplosion and wave absorption ability is. When the aluminum foam reaches the ultimate load-bearing capacity, the aluminum foam transfers the load due to compaction, resulting in stress enhancement phenomena. Through the analysis of the compression amount, floor deformation, wave dissipation capacity, and energy ratio of aluminum foam, it is concluded that the antiexplosion wave absorption effect of the sandwich structure of aluminum foam with 80%/85% group is the best; the changes of porosity and cell wall are important factors affecting the energy absorption capacity of aluminum foam.

## 1. Introduction

In recent years, various explosions have occurred frequently. Explosion containment vessels (ECVs) are widely used to completely contain the effects of explosions. Foamed aluminum (ALF) is a kind of solid material based on aluminum or aluminum alloy with three-dimensional polyhedral holes randomly distributed inside [1–4]. Because of its special porous properties, aluminum foam can have a very long and almost unchanged platform compression section during compression. While the strain increases gradually, the stress in the compression zone of the platform basically remains

unchanged and can absorb a lot of energy, so aluminum foam is an excellent energy-absorbing material [5–7], which has a good effect on reducing stress waves [8]. It has been widely used in all aspects of explosion shock protection [9–12].

The sandwich type structures, comprising the aluminum foams in the core and the thin steel cover plates, have been used to withstand higher impulse under intense dynamic events. By attaching the sandwich structure containing aluminum foam in the middle to the protective engineering or the outer surface of the building, under the impact load, the upper panel directly

acted by the external load produces a shock wave and then compresses the internal aluminum foam material; it leads to the deformation and buckling of aluminum foam, which absorbs energy, protects the bunker target, and resists the impact effect and explosion damage caused by strong dynamic load [8, 13–16]. Many scholars have studied the dynamic response and energy absorption capacity of metal foam sandwich structure, especially under explosive loading [17–21]. The results show that the dynamic properties of metal foams are affected by many factors, such as meso-structure (density, core gradation, cell wall strength, etc.), plate thickness, and loading strain rate. Functionally graded materials optimize their properties by controlling the performance gradient [22], and this method can be extended to metal foams with gradient relative density. Mortensen et al. [23, 24] made sandwich multilayer aluminum foam with core relative density  $\rho/\rho_s = 150.45\%$  and found that density classification does provide weight reduction in some strength-limited applications. The study [25] shows that the equal density foam shows the conventional behavior of metal foam. Under the condition of high strain, the foam expands to a nearly flat platform area, and then densification occurs, while the density gradient sample shows an obvious positive slope in the plateau region. As we all know, aluminum foam is a highly complex porous material. Its inherent multiscale characteristics and heterogeneity come from randomly distributed cells [26]. The meso-reaction is of great significance to the mechanical behavior of aluminum foam and the energy absorption capacity under impact loading. Li et al. [27] studied the response of gradient honeycomb sandwich structure under explosive loading by means of experiment and numerical simulation. It was found that, under the same conditions, graded sandwich plate had better antiexplosion ability than ungraded Sandwich plate. When the sandwich plate is arranged in the order of decreasing relative density, the energy absorption capacity of the sandwich plate and the attenuation effect of the contact stress on the back plate are the best. Zhou et al. [28] established a high-speed compression model of aluminum foam based on impact theory and rigid ideal plastic locking model of aluminum foam. Combined with numerical simulation, the energy absorption of gradient aluminum foam under explosive loading was studied. The results show that the larger the density gradient of aluminum foam sandwich is, the smaller the total energy input to aluminum foam is and the smaller the final deformation of aluminum foam is. Dou et al. [29] have studied the strain rate of aluminum foam sandwich panels by means of numerical simulation and think that the strain rate effect of aluminum foam becomes more and more obvious with the increase of relative density. However, the proposed meso-model of aluminum foam also has some shortcomings. Based on the three-dimensional Voronoi algorithm technology, many scholars [30–32] have established a three-dimensional meso-model of aluminum foam, which can well simulate the random distribution of aluminum foam pores and greatly

promote the study of three-dimensional meso-mechanical properties of aluminum foam. However, for the meso-model of aluminum foam established by three-dimensional Voronoi algorithm, most of the hole walls are shell elements, and their thickness is the same at any position, which obviously does not accord with the experimental results. In order to simulate the deformation and failure process of aluminum foam more truly under external load, it is very necessary to establish a wall thickness model.

## 2. Three-Dimensional Mesoscopic Modeling of Double-Layer Aluminum Foam Sandwich Panel

In this chapter, we consider the cells in the foam and give the generation algorithm of the three-dimensional model of double-layer closed-cell aluminum foam. The algorithm consists of three steps. Firstly, the random polyhedron is modeled and the geometric features are extracted; then, a three-dimensional random aluminum foam generation algorithm is developed to control the random wall thickness of the pore wall, and the aluminum foam mesh model is established by reading it into the TrueGrid software. Finally, the three-dimensional mesoscopic model of aluminum foam is imported into the LS-DYNA software to establish the explosion model of double-layer aluminum foam sandwich panel.

*2.1. Modeling Method of Three-Dimensional Meso-Model of Aluminum Foam.* In this paper, based on the three-dimensional Voronoi algorithm technology, the discrete data points are connected reasonably, the Delaunay triangulation is constructed, and the vertical bisector of the two adjacent points is connected to form the Tyson polygon [33]. The pores of the aluminum foam are regarded as the interior of the Tyson polygon, and the outline of the polygon is regarded as the cell wall of the aluminum foam. Because the wall thickness of aluminum foam is random and uneven, the random polyhedron is modeled by ABAQUS software, and the geometric features of the model are extracted. Then, based on the Voronoi algorithm, a three-dimensional foam aluminum generation algorithm with random pore size and random wall thickness was written using Python language and Fortran language. The geometric boundary parameters were read, and the random wall thickness was set to better simulate the foam aluminum under real conditions. The model uses TrueGrid's excellent grid filling technology to build a meso-grid model of aluminum foam. The finite element model of aluminum foam is obtained, as shown in Figure 1.

*2.2. Material Model and Parameters.* The bottom plate, sleeve, bracket, and end cover of the system are made of Q235 steel. The above materials and aluminum foam are numerically simulated by PLASTIC KINEMATIC material model. The model can well describe the mechanical



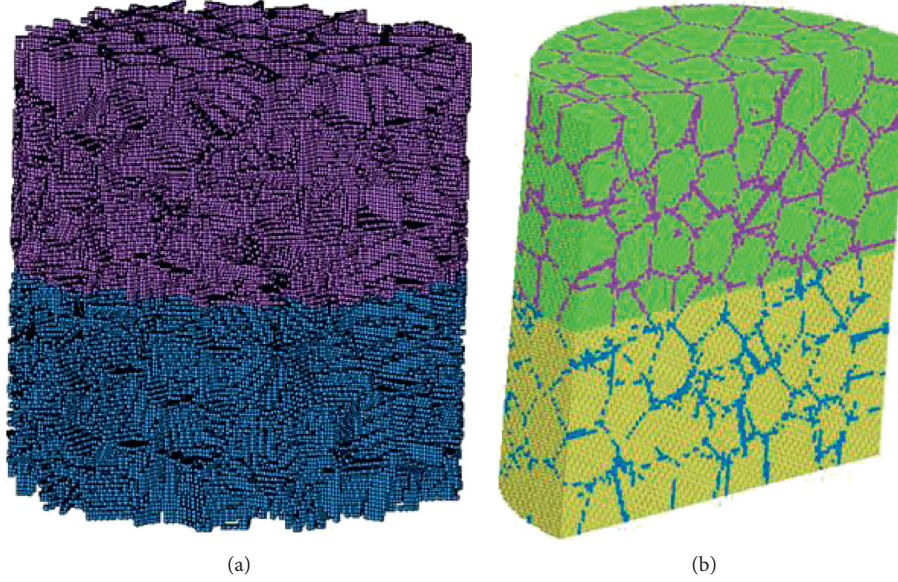


FIGURE 1: Three-dimensional microscopic view of double-layer foam aluminum.

TABLE 1: Metal material parameters.

| Metallic material | $q$ (kg/m <sup>3</sup> ) | $E$ (GPa) | $\mu$ | SIGY (MPa) | ETAN (GPa) |
|-------------------|--------------------------|-----------|-------|------------|------------|
| Q235 steel        | 7830                     | 210       | 0.274 | 235        | 6.1        |
| Aluminum foam     | 2730                     | 70        | 0.34  | 185        | 4.62       |

properties of metal materials and is widely used in numerical calculation. The calculated parameters of the metal are shown in Table 1.

The air uses the \* MAT\_NULL material model, and the equation of state is described using \* EOS\_LINEAR\_POLYNOMIAL. The expression of the equation of state is as follows:

$$p = C_0 + C_1\mu + C_2\mu^2 + C_3\mu^3 + (C_4 + C_5\mu + C_6\mu^2)E, \quad (1)$$

$$\mu = \left(\frac{1}{V}\right) - 1 = \left(\frac{\rho}{\rho_0}\right) - 1,$$

$$C_4 = C_5 = \gamma - 1.$$

Considering air as the gas in an ideal state, the coefficients of the polynomial equation are  $C_0 = C_1 = C_2 = C_3 = C_6 = 0$ . The variable coefficient  $\gamma$  is often set to 1.4, so  $C_4 = C_5 = 0.4$ .  $E_0$ ,  $\rho_0$ , and  $V_0$  are the initial energy density, the initial density, and the initial relative volume parameter values which are 1.29 g·cm<sup>-3</sup>, 0.25 MPa, and 1.0. The parameters are shown in Table 2.

The high-energy combustion explosives use \* MAT\_HIGH\_EXPLOSIVE\_BURN material model, and the equation of state uses \* EOS\_JWL to represent the pressure of the explosive product. The expression of the equation of state is shown in the following formula [18]:

TABLE 2: Material model and EOS parameters of air [20, 34, 35].

| $C_0$ (MPa) | $C_1$ | $C_2$ | $C_3$ | $C_4$ | $C_5$ | $C_6$ | $\rho$ (kg/m <sup>3</sup> ) |
|-------------|-------|-------|-------|-------|-------|-------|-----------------------------|
| -0.1        | 0     | 0     | 0     | 0.4   | 0.4   | 0     | 1.225                       |

$$p = A \left(1 - \left(\frac{\omega}{R_1 V}\right)\right) e^{-R_1 V} + B \left(1 - \left(\frac{\omega}{R_2 V}\right)\right) e^{-R_2 V} + \left(\frac{\omega E}{V}\right), \quad (2)$$

where  $E$  and  $V$  are energy density and relative volume, respectively. When the initial calculation is performed, they should be given initial values  $E_0$  and  $V_0$ ;  $A$ ,  $B$ , and  $E_0$  are pressure units;  $R_1$ ,  $R_2$ , OMEG, and  $V_0$  are dimensionless. The specific material parameters are shown in Table 3.

The outer boundary of the air adopts a nonreflective boundary, the bottom of the bottom plate is set as a fixed boundary, and the contact setting is set: the cover plate is bound to connect with the bottom plate, \* CONTACT\_TIED\_SURFACE\_TO\_SURFACE, cover plate is connected with aluminum foam, and the aluminum foam is automatically contacted with the bottom plate by automatic surface contact \* CONTACT\_AUTOMATIC\_SURFACE\_TO\_SURFACE.

**2.3. Modeling of Double-Layer Aluminum Foam Sandwich Panel.** The energy absorption effect and explosion resistance of aluminum foam are closely related to the density, and the

TABLE 3: Material model parameters and JWL parameters of TNT.

| $\rho$ (kg/m <sup>3</sup> ) | $A$ (GPa) | $B$ (GPa) | $R_1$ | $R_2$ | PCJ (GPa) | $E$ (J/m <sup>3</sup> ) | $D$ (m/s) | $\nu$ | $\Omega$ |
|-----------------------------|-----------|-----------|-------|-------|-----------|-------------------------|-----------|-------|----------|
| 1500                        | 347.6     | 3.39      | 4.15  | 0.95  | 17.92     | 6.34e9                  | 6957.2    | 1.0   | 0.28     |

difference of porosity will inevitably lead to the change of the density of aluminum foam. The arrangement of different porosity of multilayer aluminum foam will also have a significant impact on the dynamic response of aluminum foam sandwich panels. Six groups of different structures were obtained by permutation and combination of aluminum foam with three densities, namely, 80%, 80%, 85%, 85%, 90%, 90%, 90%, 90%, 80%, and 90%, respectively. The numerical model is shown in Figure 2. In the model, the mass of the charge is fixed as 520 g, the density is 1.601 g, the radius is 0.086 m, and the spherical charge and the height of the charge are fixed as 25 cm. The thickness of cover plate is fixed at 10 mm, the total height of aluminum foam sandwich is fixed at 78 mm, and the height of single layer is 39 mm. As shown in Figure 2, the bottom plate support of the model is set as a fixed constraint, and a downward prestressed load is applied on the upper surface to clamp the aluminum foam. The pore sizes  $R_{\min}$  and  $R_{\max}$  in the model are 1 mm and 3 mm, respectively, and the average pore size is 2 mm. In order to ensure the calculation accuracy of the flow field in the container, the element size in the finite element model is 0.2 mm.

**2.4. Layout of Measuring Points.** After the explosion of the explosive, the explosion products and shock waves of high temperature and high pressure will be produced, which will have a strong destructive effect on the surrounding structures. A measuring point is set up in the sandwich structure bearing the explosion shock wave to test the strength of the stress wave. The location of the measuring points is shown in Figure 3, and three groups of stress measuring points are set. The stress wave intensities of the cover plate were transmitted to the first layer of aluminum foam, the first layer of aluminum foam to the second layer of aluminum foam and the second layer of aluminum foam to the bottom plate are measured, respectively, and the test results are recorded as  $\sigma_1$ ,  $\sigma_2$ , and  $\sigma_3$ , respectively.  $\sigma_1$  is the mean stress of the uppermost element of the first layer of aluminum foam,  $\sigma_2$  is the average stress of the lowest element of the first layer and the uppermost element stress of the second layer, and  $\sigma_3$  is the mean stress of the uppermost element of the bottom plate.

### 3. Validation

**3.1. Experiment Setup.** In this paper, the explosion test device shown in Figure 4 is designed to study the anti-explosion performance of double-layer aluminum foam sandwich structure. The shock wave produced by the explosion of the spherical TNT charge causes the cover plate to move and deform and compress the aluminum foam under the explosion load. In the experiment, the blasting height of the charge is fixed at 25 cm, the thickness of the upper panel

is fixed at 10 mm, the height of each layer of aluminum foam is 39 mm, and the total thickness of the aluminum foam sandwich is fixed as 78 mm.

In order to evaluate the explosion resistance of different aluminum foam sandwich structures, a cylindrical pressure sensor is installed on the upper surface of the base plate. According to the fixed mode of the whole device, strain gauges are set at the positions of S1, S2, and S3 on the lower surface of the base plate, which mainly test the dynamic response of the bottom plate under the explosion impact load. As shown in Figure 5, the load and deformation of the bottom plate are monitored, respectively. The pressure sensor is in the center of the upper surface of the base plate. The strain gauge S1 is in the center of the lower surface of the base plate.

**3.2. Comparative Analysis with the Test.** In order to verify the correctness of the model, the strain test values of 80%/90% group and 80%/85% group of aluminum foam were randomly selected and compared. Figure 6 is the comparison diagram of the strain time history curve of 80%/90% foam aluminum test value and simulation value of each measuring point. S1 (2, 3)-e represents the experimental value, and S1 (2, 3) represents the simulation value. As can be seen from the chart, the strain peak value of each measuring point has little difference, and the trend of the strain curve waveform is the same, which tends to be smooth with the increase of time. The test simulation results verify the test results. Figure 7 shows the test height and simulation height of 80% ram 85% aluminum foam after compression. The initial height of double-layer aluminum foam is 78 mm. After the test, the height of double-layer aluminum foam becomes 46.55 mm, and the compression amount is 31.45 mm. The compression amount of the numerical simulation is 32.8 mm, and the difference between the experimental value and the simulation value is small, which can verify the correctness of the model.

## 4. Analysis of Energy Absorption Mechanism of Aluminum Foam

**4.1. Analysis of the Movement Process of Double-Layer Foam Aluminum.** This section explores the motion process of 85%/90% foam aluminum sandwich panels under explosive loading.

In the first stage, the explosive center detonates until the shock wave reaches the cover plate and begins to compress the aluminum foam. As shown in the figure, during 0 ms, the explosive initiates at a single point center, and the shock wave front begins to expand outward with the sphere as the center, reaching the cover plate at a certain time between 0.07 ms and 0.08 ms, and the cover plate obtains a certain speed to compress the aluminum foam downward, as shown

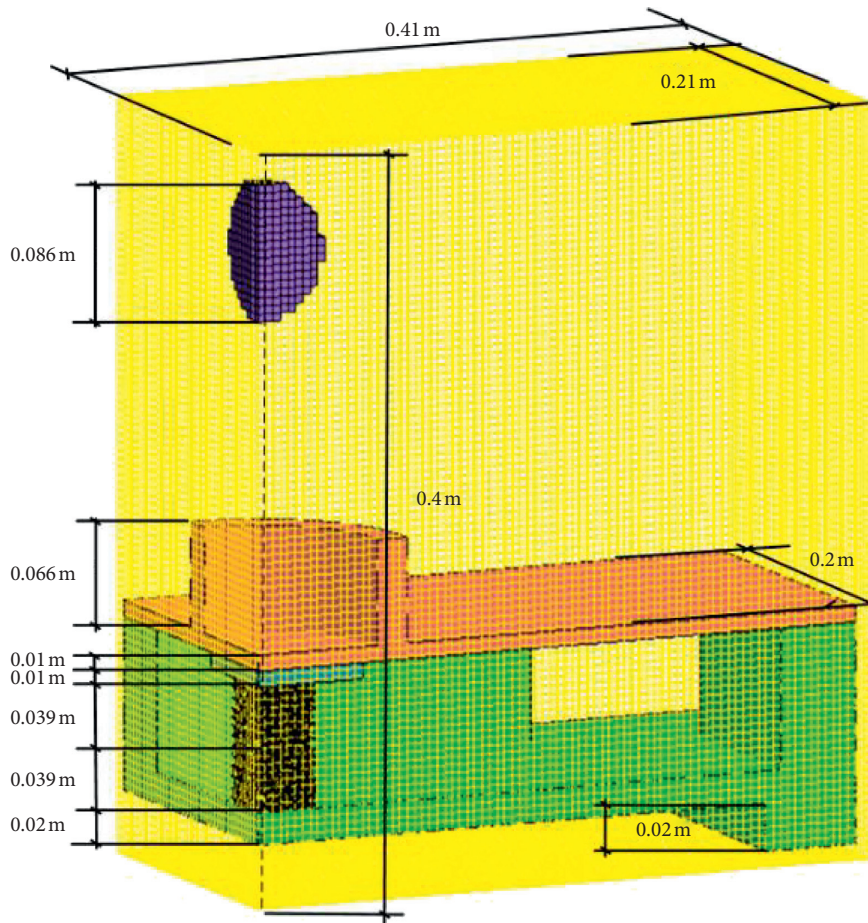


FIGURE 2: Double-layer aluminum foam sandwich (AFS) structure.

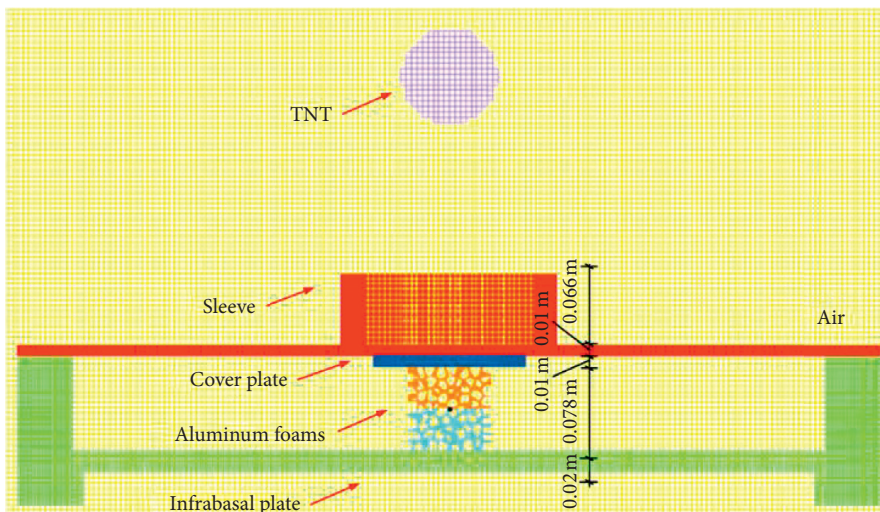


FIGURE 3: Stress measuring point arrangement.

in the pressure cloud diagram of aluminum foam on the right side of Figure 8.

In the second stage, the cover plate begins to compress aluminum foam to fully compress double-layer aluminum foam. As shown in Figure 9, contact compression occurs at the delamination interface at 0.08 ms, and the stress wave

begins to travel from the top foam to the bottom foam, resulting in reflected and transmitted waves. Because the wave impedance of 90% porosity aluminum foam is less than 80% porosity aluminum foam, and 90% porosity is high, so 90% porosity aluminum foam absorbs most of the energy and produces great deformation. The stress wave is

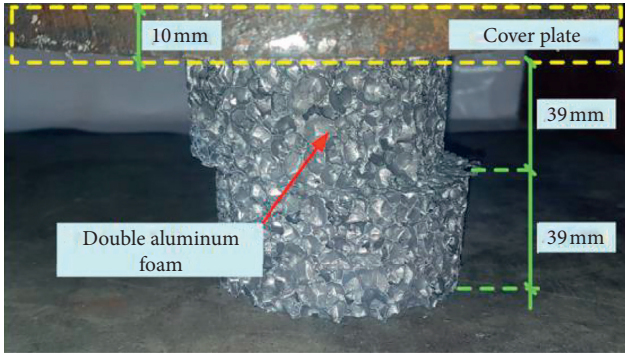


FIGURE 4: Sandwich plate antiexplosion test device.

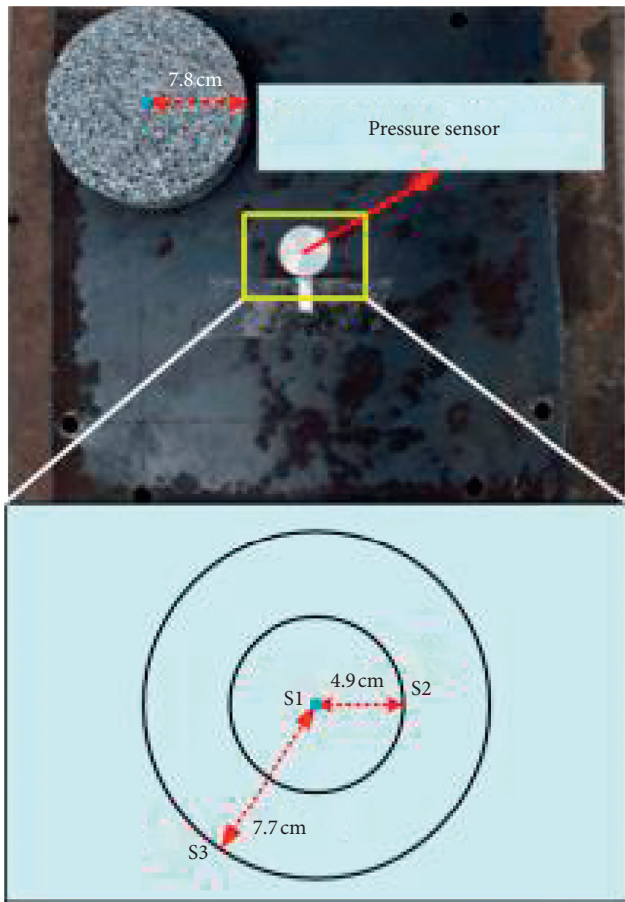


FIGURE 5: Sensor settings.

transmitted to the bottom plate at 0.24 ms and begins to act to the bottom plate, so that the floor begins to compress and deform. During the explosion process to 1.0 ms, we can see that 90% aluminum foam is completely compressed almost at some time between 1.0 ms and 1.24 ms, while, after 1.0 ms, 80% aluminum foam is compressed as the main energy-absorbing material. This is very helpful for us to study the energy absorption mechanism of layered aluminum foam.

From the summary of the previous research background, we can see that the stress-strain performance of aluminum foam is different under the sorting condition of different aluminum foam density, but the simulation analysis of the

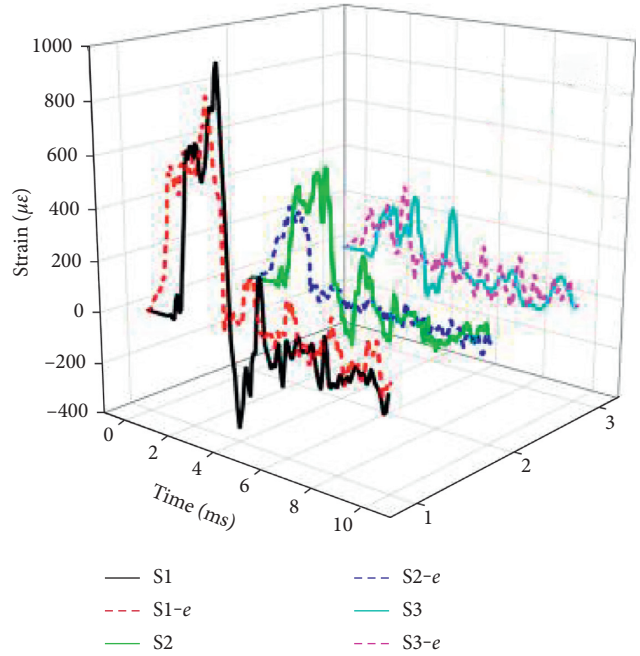


FIGURE 6: 80%/90% foam aluminum strain diagram.

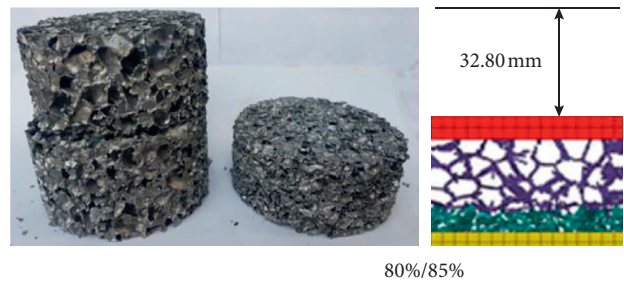


FIGURE 7: 80%/85% aluminum foam compression.

ordering of density gradient is still less. In order to further analyze the effect of density gradient on the mechanical properties of closed-cell aluminum foam, this section is based on a three-dimensional meso-model to study the explosion resistance of aluminum foam specimens under different density order and the energy absorption characteristics of aluminum foam with different density gradient under explosive impact loading.

*4.2. Analysis of Delamination Characteristics of Double-Layer Aluminum Foam.* You can see significant compression when the time reaches 0.48 ms. When compressed to 1.24 ms, the compression of aluminum foam is completed. Figure 10 shows the deformation distribution of aluminum foam under explosive loading under six groups of porosity combinations during 1.24 ms. It can be seen from the figure that the deformation of aluminum foam in 90%/80% group is the largest, with a deformation amount of 34.23 mm. The deformation of the pore wall of aluminum foam is mainly concentrated in the area with large porosity. Through the

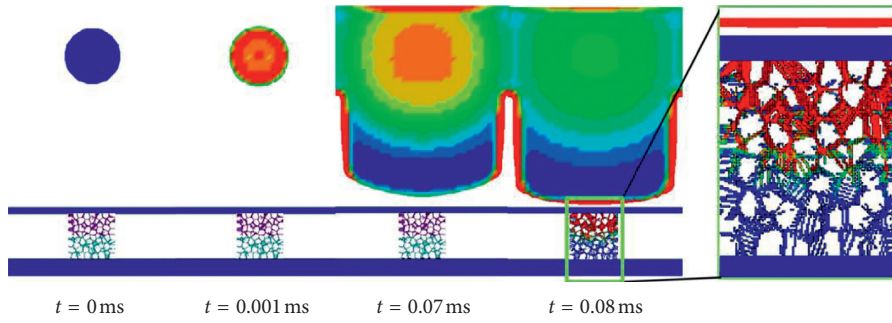


FIGURE 8: Compressed foam aluminum cover.

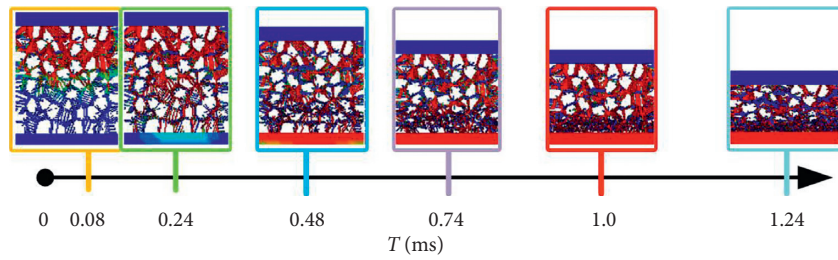


FIGURE 9: Compressed foam aluminum cover.

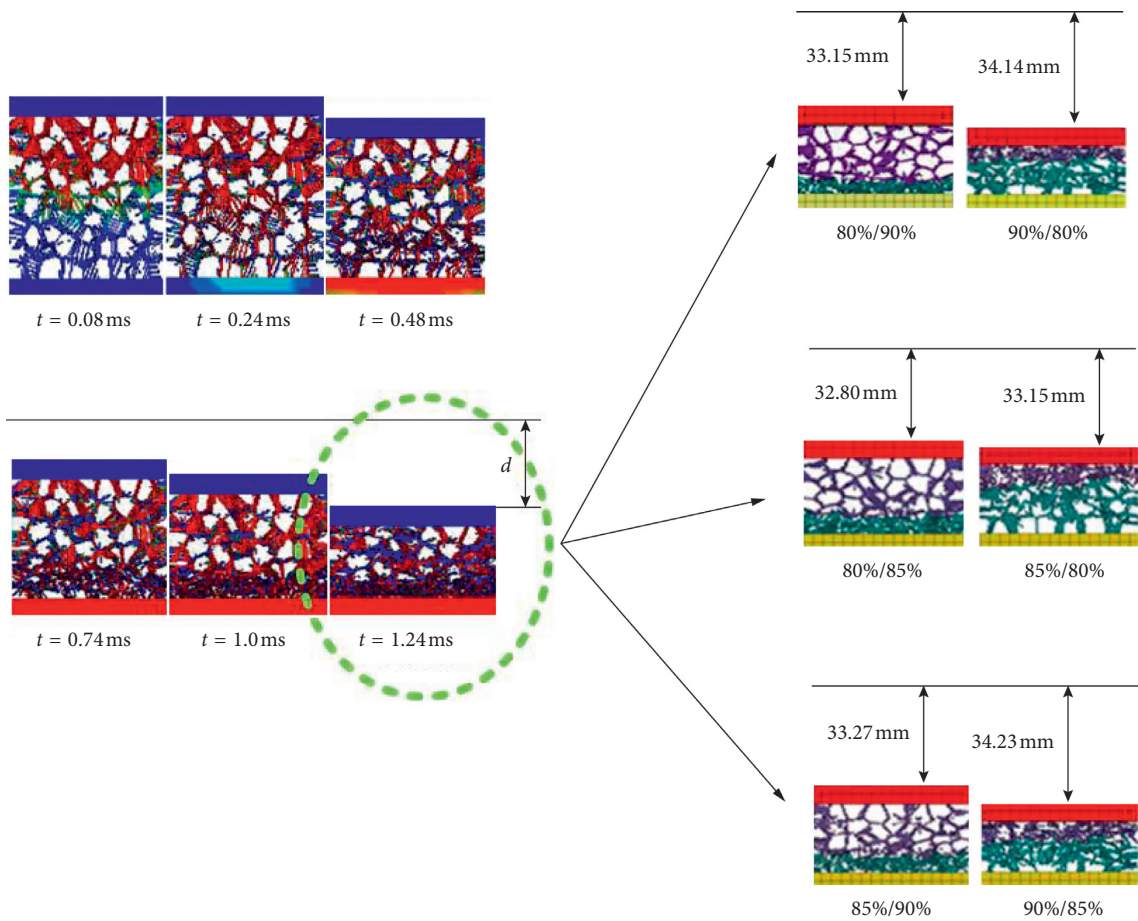


FIGURE 10: Deformation analysis of six groups of structures.

comparison of the structures of each group, it is found that if the porosity of the upper layer aluminum foam is higher than the lower layer aluminum foam, the compression amount of the double-layer aluminum foam sandwich structure is larger, such as group 90%/80% and group 90%/85%. When the porosity of the upper layer aluminum foam is small, the two layers of aluminum foam begin to deform at the same time, but when the porosity of the upper layer aluminum foam is larger, it will not begin to deform at the same time. Before the upper layer of aluminum foam is compressed to a certain extent, the contact stress between the two layers of aluminum foam is greater than the yield stress of the lower layer of aluminum foam, and the lower layer of aluminum foam begins to deform.

As can be seen from Table 4, the maximum peak displacement of the 90%/80% group is the largest among the six groups, followed by the 90%/85% group displacement. The order of the bottom plate displacement under each combination is as follows: 80%/90% < 85%/90% < 80%/80% < 90%/85% < 90%/85% < 90%/80%. Under the only combination of the two groups of porosity, the large upper porosity will increase the compression of the sandwich structure of double-layer aluminum foam and reduce the displacement of the bottom plate. With the increase of time, the elastic deformation of the bottom plate ends and enters the plastic deformation stage, and the deformation of the bottom plate tends to be stable. After reaching stability, the minimum displacement of the bottom plate is 80%/90% group, followed by 85%/90% group. The order of displacement is as follows: 80%/85% < 85%/90% < 80%/80% < 90%/80% < 90%/85% < 90%/80%. Since the bottom plate displacement of the aluminum foam in the 80%/85% group is the smallest after the stress balance, it is considered that the reorganization is the total optimal group of the six groups. It can be found that, under the positive sequence of porosity, the displacement of the bottom plate is smaller. By comparing the compression amount of aluminum foam, it can be found that the compression amount of aluminum foam in the group of 90%/85% is the largest, but the deformation of the bottom plate is not the smallest, but very large. This also shows that it is not that the greater the compression amount of aluminum foam is, the better the antiexplosion and wave-absorbing ability is and the better the protective effect on the substructure is. The analysis shows that after the aluminum foam reaches the ultimate bearing capacity, the aluminum foam produces stress enhancement due to compaction, so the compression amount of aluminum foam is the largest, but its wave absorption capacity is not the best.

In order to quantitatively describe compressive quantities of double-layer aluminum foams and compare compressive quantities of single-layer aluminum foams, we draw a curve of displacement time history of different porosity combinations in Figure 11. It can be seen from graphs that foam aluminum reaches maximum displacement rapidly during elastic stage and then decreases and eventually tends to be smooth. Results show that compressive strength of double-layer foamed aluminum bottom plate is smaller than that of single-layer aluminum foam plate. Design of double-layer aluminum foam can improve antiexplosion property of

TABLE 4: Displacement of double-layer foamed aluminum under each combination.

|                               | 80/90 | 90/80 | 80/85 | 85/80 | 90/85 | 85/90 |
|-------------------------------|-------|-------|-------|-------|-------|-------|
| Maximum deformation           | -1.57 | -2.01 | -1.76 | -1.84 | -1.94 | -1.61 |
| Deformation after equilibrium | -1.00 | -1.55 | -0.28 | -1.17 | -1.51 | -0.54 |

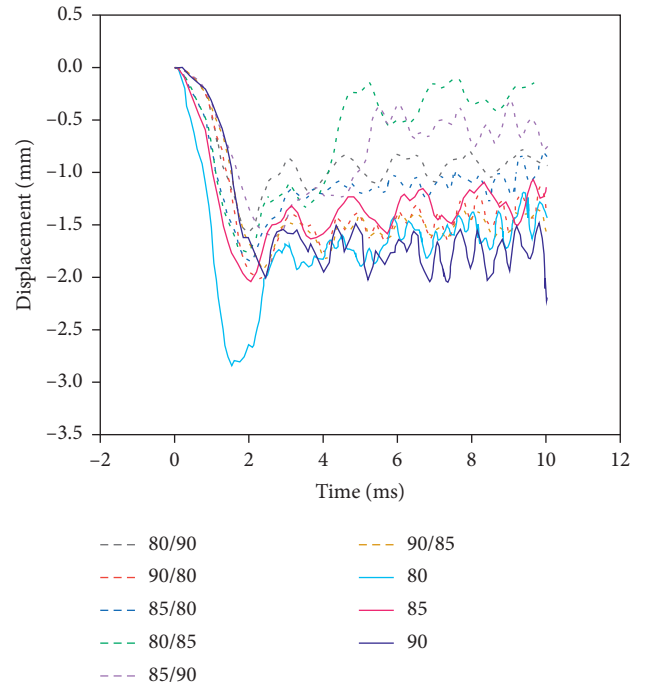


FIGURE 11: Displacement of single-layer aluminum foam bottom plate.

TABLE 5: Upper aluminum foam energy.

|                     | 80/90  | 80/85  | 90/80  | 90/85  | 85/80  | 85/90  |
|---------------------|--------|--------|--------|--------|--------|--------|
| Total energy (J)    | 688.11 | 739.35 | 584.55 | 503.74 | 970.24 | 898.67 |
| Internal energy (J) | 604.19 | 665.91 | 556.76 | 467.01 | 952.9  | 846.21 |
| Kinetic energy (J)  | 83.92  | 73.44  | 27.79  | 36.73  | 17.34  | 52.46  |

TABLE 6: Energy of aluminum foam in the lower layer.

|                     | 80/90   | 80/85   | 90/80   | 90/85   | 85/80   | 85/90   |
|---------------------|---------|---------|---------|---------|---------|---------|
| Total energy (J)    | 1125.98 | 1734.18 | 1322.07 | 1125.98 | 1154.61 | 1605.32 |
| Internal energy (J) | 1064.42 | 1096.89 | 1240.25 | 1031.1  | 1085.62 | 1477.88 |
| Kinetic energy (J)  | 61.56   | 637.29  | 81.82   | 94.88   | 68.99   | 127.44  |

TABLE 7: Floor energy.

|                     | 80/90  | 80/85  | 90/80  | 90/85  | 85/80  | 85/90  |
|---------------------|--------|--------|--------|--------|--------|--------|
| Total energy (J)    | 126.12 | 121.55 | 249.56 | 191.24 | 162.37 | 176.41 |
| Internal energy (J) | 83.14  | 102.59 | 187.11 | 141.62 | 125.78 | 133.26 |
| Kinetic energy (J)  | 42.98  | 18.96  | 62.45  | 49.62  | 36.59  | 43.15  |

aluminum foam effectively, eliminate partial explosion shock wave effectively, and play positive role in protecting lower structure.

**4.3. Energy Analysis.** Tables 5–7 provide energy for upper and lower aluminum foams and bottom plates, respectively. According to the tables, the total energy of lower aluminum foam is higher than that of upper aluminum foam, which is consistent with wave dissipation rule of aluminum foam at upper and lower level. The fluctuation dissipation rate of low aluminum foam is obviously higher than that of aluminum foam. As shown in Table 5, the difference between total energy and internal energy is smaller than that of upper foam aluminum, and the kinetic energy of 80%/90% group is maximum, which is consistent with minimum law of peak displacement of bottom plate of upper foam aluminum group, followed by group 80%/85%. As shown in Table 6, the gap between total energy and internal energy is larger during lower aluminum foams. The maximum kinetic energy is in group 80%/85%, and the minimum kinetic energy is in group 80%/90%. As shown in Table 7, total energy, internal energy, and kinetic energy are minimal in group 80%/85%. The results show that porosity has great influence on energy absorption characteristics of samples; when the total energy of aluminum foam increases, the energy of bottom plate decreases. There are mainly two types of aluminum foam to improve porosity; one case is pore quantity that increases; another case is pore wall thickness that decreases. Therefore, more cell walls produce larger deformation plastic strain absorbing more energy under the same strain condition. However, due to low porosity, the overall antiblast performance of this device is better but critical point of complete compression and incomplete compression can be found which ensures that load can be absorbed completely without destroying floor.

Figure 12 shows the ratio of internal energy to kinetic energy of the floor arranged at all levels of density. It can be seen from the diagram that the energy ratio of the floor fluctuates with the increase of time, indicating that the kinetic energy ratio of the floor continues to decrease and the kinetic energy is converted into internal energy and finally tends to be stable. The soleplate of 80%/85% group showed the maximum peak of energy ratio during 4.496 ms, and obvious secondary damage occurred at the interface, resulting in secondary work done by aluminum foam, so it consumes a lot of energy. Since the displacement of the bottom plate of the 80%/85% group is the smallest, it can be concluded that this group is the best group in density order.

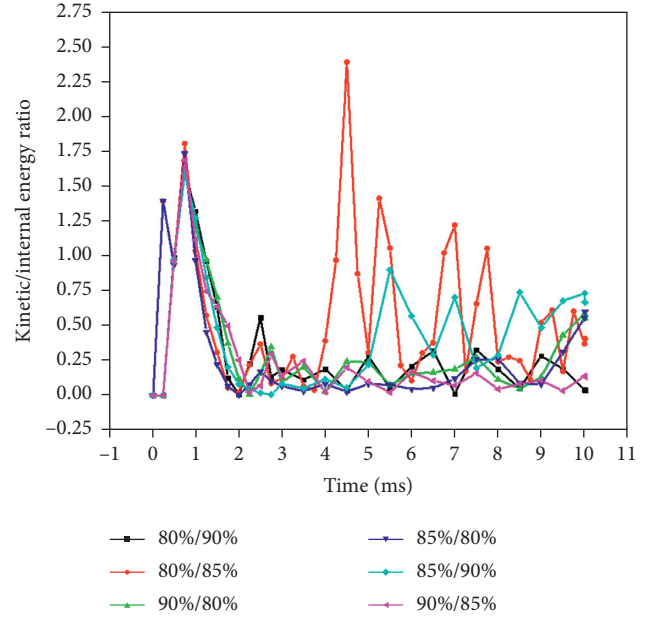


FIGURE 12: Energy ratio of aluminum foam.

## 5. Investigation on Three-Dimensional Meso-Mechanical Properties of Double-Layer Aluminum Foam Sandwich Panels

In order to analyze the relationship between meso-parameters and mechanical properties of materials under explosive loading, we study the energy absorption characteristics of aluminum foam materials and the effects of meso-failure mechanism of aluminum foam cell wall on its antiexplosion performance and energy absorption. In this section, the dynamic mechanical properties of aluminum foam specimens are studied by simulating the double-layer aluminum foam sandwich panel under explosive load.

The literature research [36] shows that the stress wave propagates from one medium to another; if the wave impedance of the two media is different, then the stress wave will be reflected and transmitted at the interface of the two media, and its intensity satisfies the following relationship:

$$\left(\frac{\sigma_R}{\sigma_I}\right) = \frac{(\rho_2 C_2 - \rho_1 C_1)}{(\rho_2 C_2 + \rho_1 C_1)}, \quad (3)$$

$$\left(\frac{\sigma_T}{\sigma_I}\right) = \frac{(2\rho_2 C_2)}{(\rho_2 C_2 + \rho_1 C_1)}. \quad (4)$$

Among them,  $\sigma_I$ ,  $\sigma_R$ , and  $\sigma_T$  represent the incident wave, reflected wave, and transmitted wave intensity of the shock wave at the interface, respectively;  $\rho_1 C_1$  represents the wave impedance of the first layer medium, and  $\rho_2 C_2$  represents the wave impedance of the second layer medium. It can be seen from formulas (3) and (4) that when the aluminum foam sandwich is delaminated in density, due to the difference of medium wave impedance on both sides of the interface, the shock wave that continues to propagate along the foam

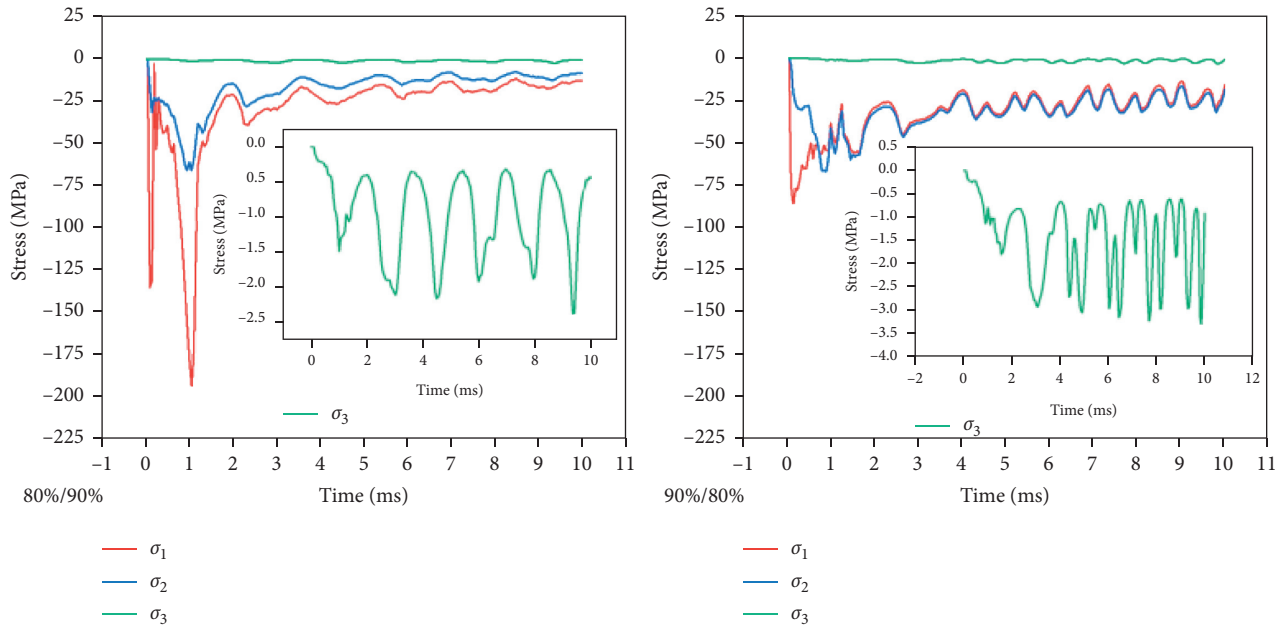


FIGURE 13: Double-layer aluminum foam with porosity of 80% and 90%.

aluminum to the bottom plate after transmission through the interface decreases, while the shock wave reflected from the interface near the explosion source end may further compress the deformed area of the aluminum foam, so delamination of the aluminum foam may increase the energy absorption value of the sandwich structure core of the foam aluminum sandwich. At the same time, the load acting on the protected structure is reduced. Therefore, we established a three-dimensional mesoscopic model of double-layer aluminum foam and quantitatively analyzed the attenuation effect of stress wave strength of double-layer aluminum foam sandwich structure.

When the porosity combination of aluminum foam is 80% and 90%, the distribution of stress waves in the sandwich structure is shown in Figure 13. Intuitively, in Figure 13, the stress value of the first layer reached 135.94 MPa at 0.102 ms and reached the second peak at 1.057 ms due to emission, indicating that, in the steel plate, the change rate of stress wave is extremely high, the stress wave intensity of the second layer is 31.17 MPa, and the stress wave intensity of the third layer is 2.38 MPa. The attenuation of the explosion shock wave is as high as 77.07% after passing through the first layer of aluminum foam, the attenuation of the explosion wave after passing through the second layer of aluminum foam reaches 92.36%, and the overall wave dissipation rate is 98.25%. The rising edge of the stress wave in the first layer is  $10.2 \mu\text{s}$ , the rising edge of the stress wave in the second layer is  $14.9 \mu\text{s}$ , and the rising edge of the stress wave in the third layer can reach 0.9 ms. After aluminum foam, the rising time of stress wave has been greatly improved. When the porosity of the upper layer of aluminum foam is 90% and the porosity of the lower layer of aluminum foam is 80%, the stress value of the first layer is 86.70 MPa, the attenuation of explosion shock wave after the first layer of aluminum foam is 64.46%, the stress value of explosion

wave after the second layer of aluminum foam is 2.94 MPa, the attenuation rate is 90.45%, and the overall wave dissipation rate is 96.61%. It can be found that whether it is the first layer of aluminum foam or the second layer of aluminum foam, the positive sequence arrangement of porosity has higher wave dissipation rate and the phenomenon of interface delamination is more obvious.

When the porosity combination of aluminum foam is 80% and 85%, the distribution of stress waves in the sandwich structure is shown in Figure 14. When the upper porosity of the aluminum foam is 80% and the porosity of the lower aluminum foam is 85%, the stress value of the first layer of the explosion shock wave reaches 158.76 MPa, and the attenuation is as high as 59.07% after passing through the first layer of aluminum foam; the stress value of the explosion wave after passing through the second layer of aluminum foam is 2.38 MPa, the attenuation rate is 95.80%, and the overall wave dissipation rate is 98.51%. When the porosity of the upper layer of aluminum foam is 85% and that of the lower layer is 80%, the stress value of the first layer is 151.01 MPa, and the attenuation rate of the explosion shock wave after the first layer of aluminum foam is 49.02%. After the second layer of aluminum foam, the stress value of the explosion wave is 2.27 MPa, the attenuation rate is 95.34%, and the overall wave dissipation rate is 98.49%. It is also found that whether it is the first layer of aluminum foam or the second layer of aluminum foam, the positive sequence arrangement of porosity has a higher wave dissipation rate and the phenomenon of interface delamination is more obvious.

When the porosity combination of aluminum foam is 85% and 90%, the distribution of stress waves in the sandwich structure is shown in Figure 15. When the porosity of the upper layer of aluminum foam is 85% and that of the lower layer is 90%, the stress value of the first layer is



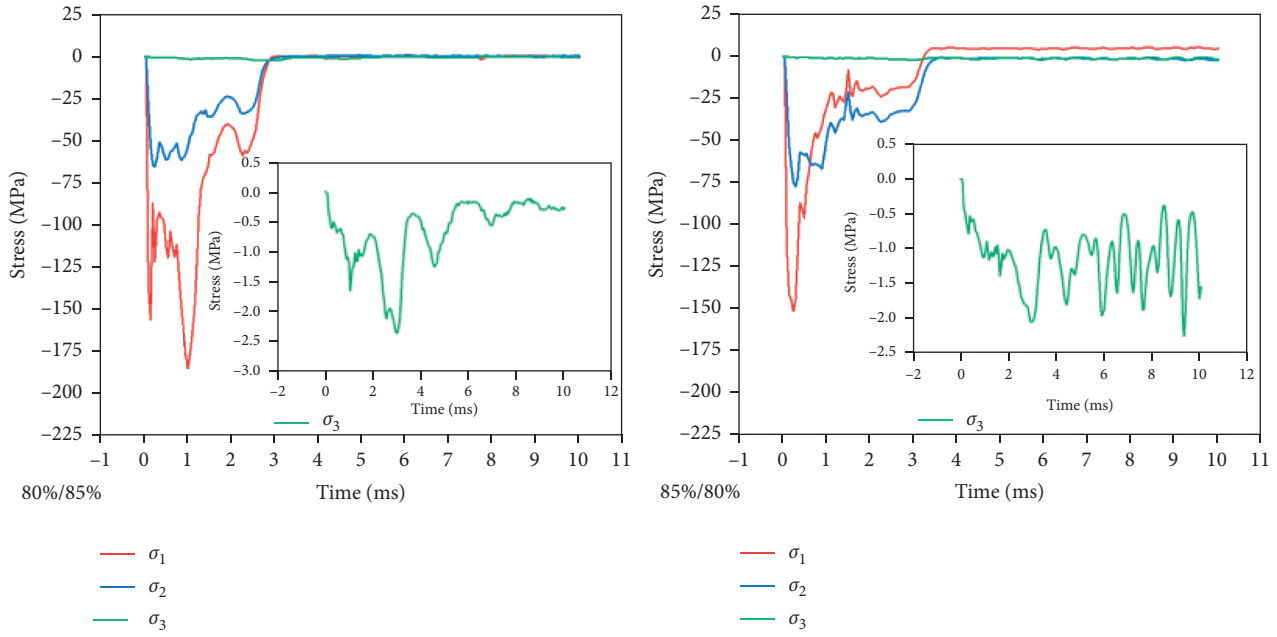


FIGURE 14: Double-layer aluminum foam with porosity of 80% and 85%.

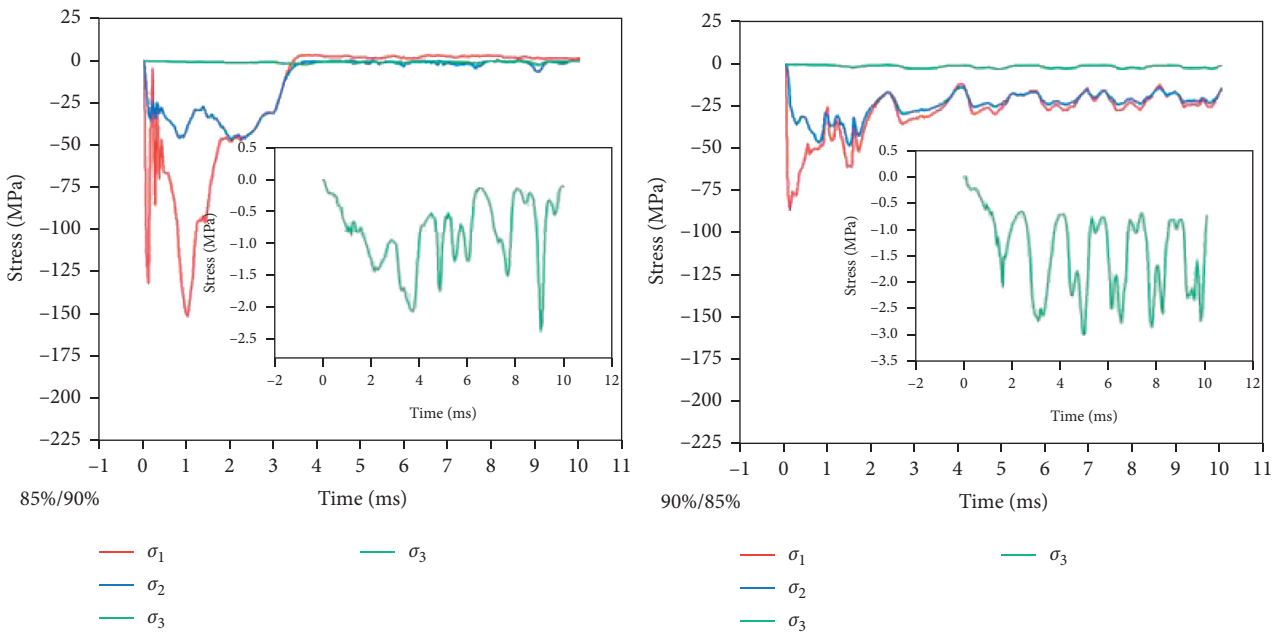


FIGURE 15: Double-layer aluminum foam with porosity of 85% and 90%.

129.36 MPa. The shock wave of the explosion after the first layer of aluminum foam attenuates by 71.44%. After the second layer of aluminum foam, the stress value of the explosion wave is 2.013 MPa, the attenuation rate is 96.81%, and the overall wave dissipation rate is 98.43%. When the porosity of the upper layer of aluminum foam is 90% and the porosity of the lower layer of aluminum foam is 85%, the stress value of the first layer is 86.63 MPa; the explosion shock wave attenuates 59.15% after the first layer of aluminum foam, and the stress value of the explosion wave after

the second layer of aluminum foam is 2.72 MPa; the attenuation rate is 92.31%, respectively, and the overall wave dissipation rate is 96.86%. It is also found that whether it is the first layer of aluminum foam or the second layer of aluminum foam, the positive sequence arrangement of porosity has a higher wave dissipation rate and the phenomenon of interface delamination is more obvious.

Due to different porosity of aluminum foam, different reflection and refraction phenomena occur after explosion wave propagation, so there is stress difference between upper

surfaces of aluminum foam layer. It can be concluded from the above picture that the overall wave dissipation rates of six groups of aluminum foams with different porosity have little difference. By comparison, it is found that the porosity is arranged in positive order, and the difference between the upper and lower layers is small, and the wave dissipation rate of the lower layer aluminum foam is higher. To sum up, the antiexplosion and wave-absorbing ability of the double-layer foam sandwich structure of 80%/85% group is the best, and the greater the porosity is, the better the antiexplosion ability is.

## 6. Conclusions

In this paper, a three-dimensional mesoscopic model of double-layer aluminum foam sandwich panel is established, and the fluid-solid coupling method is used to analyze the flow field motion of explosion shock wave of aluminum foam with different porosity. A series of conclusions are obtained through numerical simulation and analysis:

- (1) On the basis of Voronoi algorithm, the aluminum foam generation algorithm with random three-dimensional pore size and wall thickness is compiled by using Python language and Fortran language. The three-dimensional mesoscopic model of double-layer closed-cell aluminum foam sandwich panel is established by using LS-DYNA and ABAQUS software, and the random wall thickness is set to better simulate the real aluminum foam model.
- (2) The strain test values of 80%/90% group and 80%/85% group of aluminum foam were randomly selected and compared with the simulation results. The experimental results are consistent with the simulation results, which verifies the correctness of the three-dimensional meso-model.
- (3) Six kinds of three-dimensional mesoscopic models of explosive loading of double-layer aluminum foam with different porosity arrangement were established in LS-DYNA. The numerical simulation results show that, in the group with the same porosity combination, when the porosity of the upper layer aluminum foam is greater than that of the lower layer aluminum foam, the compression amount of the sandwich structure of double-layer aluminum foam is larger, but the displacement of the bottom plate is small; it is not that the greater the compression amount of aluminum foam, the better the anti-explosion and wave absorption capacity. When the aluminum foam reaches the ultimate bearing capacity, the foam aluminum produces stress enhancement due to compaction.
- (4) Through the mesoscopic simulation of aluminum foam, it is found that the changes of porosity are very important factors affecting the energy absorption capacity of aluminum foam.
- (5) Through the analysis of the compression amount, floor deformation, wave dissipation capacity, and energy ratio of aluminum foam, it is concluded that

the antiexplosion wave absorption effect of the sandwich structure of aluminum foam with 80%/85% group is the best.

## Data Availability

The drawings and tables in the article are all original data, which are all obtained by experiments and computer simulations. In the future, when readers need the data in the article for secondary development, they can e-mail the authors to provide it. The drawings and tables in the article can be edited without any problems.

## Conflicts of Interest

The authors declare that they have no conflicts of interest regarding their work.

## Acknowledgments

The financial support from the Science and Technology Innovation Project (KYGYZB0019003) and the Department of Infrastructure Barracks (China) is gratefully acknowledged.

## References

- [1] I. Duarte, M. Vesenjaj, and L. Krstulović-Opara, "Compressive behaviour of unconstrained and constrained integral-skin closed-cell aluminium foam," *Composite Structures*, vol. 154, pp. 231–238, 2016.
- [2] J. H. Kim, D. Kim, M.-G. Lee, J. K. Lee, and M.-G. Lee, "Multiscale analysis of open-cell aluminum foam for impact energy absorption," *Journal of Materials Engineering and Performance*, vol. 25, no. 9, pp. 3977–3984, 2016.
- [3] B. Hamidi Ghaleh Jigh, H. Hosseini Toudeshky, and M. A. Farsi, "Experimental and multi-scale analyses of open-celled aluminum foam with hole under compressive quasi-static loading," *Journal of Alloys and Compounds*, vol. 695, pp. 133–141, 2017.
- [4] M. Vesenjaj, C. Veyhl, and T. Fiedler, "Analysis of anisotropy and strain rate sensitivity of open-cell metal foam," *Materials Science and Engineering: A*, vol. 541, pp. 105–109, 2012.
- [5] G. Lu and T. X. Yu, *Energy Absorption of Structures and Materials*, Woodhead Publishing Ltd., Cambridge, MA, USA, 2003.
- [6] S. Cheng, X. Zhao, B. Xiao, and Y. Xin, "Compression tests on aluminum honeycomb and epoxy resin Sandwich panels," *Emerging Materials Research*, vol. 4, no. 2, pp. 157–164, 2015.
- [7] D. K. Rajak, L. A. Kumaraswamidhas, S. Das, and S. Senthil Kumaran, "Characterization and analysis of compression load behaviour of aluminium alloy foam under the diverse strain rate," *Journal of Alloys and Compounds*, vol. 656, pp. 218–225, 2016.
- [8] R. P. Merrett, G. S. Langdon, and M. D. Theobald, "The blast and impact loading of aluminium foam," *Materials & Design*, vol. 44, pp. 311–319, 2013.
- [9] J. Park and H.-J. Choi, "Experiments and numerical analyses of HB400 and aluminum foam sandwich structure under landmine explosion," *Composite Structures*, vol. 134, pp. 726–739, 2015.

- [10] M. F. Ashby, A. G. Evans, N. A. Fleck, L. J. Gibson, and J. W. Hutchinson, *Metal Foams: A Design Guide*, Butterworth-Heinemann, Oxford, UK, 2000.
- [11] A. G. Evans, J. W. Hutchinson, and M. F. Ashby, "Multi-functionality of cellular metal systems," *Progress in Materials Science*, vol. 43, no. 3, pp. 171–221, 1998.
- [12] M. V. Hosur, U. K. Vaidya, C. Ulven, and S. Jeelani, "Performance of stitched/unstitched woven carbon/epoxy composites under high velocity impact loading," *Composite Structures*, vol. 64, no. 3-4, pp. 455–466, 2004.
- [13] G. S. Langdon, D. Karagiozova, M. D. Theobald et al., "Fracture of aluminium foam core sacrificial cladding subjected to air blast loading," *International Journal of Impact Engineering*, vol. 37, pp. 638–651, 2018.
- [14] G. W. Ma and Z. Q. Ye, "Energy absorption of double-layer foam cladding for blast alleviation," *International Journal of Impact Engineering*, vol. 34, no. 2, pp. 329–347, 2007.
- [15] G. S. Langdon, D. Karagiozova, M. D. Theobald et al., "Fracture of aluminium foam core sacrificial cladding subjected to air-blast loading," *International Journal of Impact Engineering*, vol. 37, no. 6, pp. 638–651, 2010.
- [16] J. A. Main and G. A. Gazonas, "Uniaxial crushing of sandwich plates under air blast: influence of mass distribution," *International Journal of Solids and Structures*, vol. 45, no. 7-8, pp. 2297–2321, 2008.
- [17] M. Peroni, G. Solomos, and V. Pizzinato, "Impact behaviour testing of aluminium foam," *International Journal of Impact Engineering*, vol. 53, pp. 74–83, 2013.
- [18] L. Jing, Z. Wang, and L. Zhao, "Dynamic response of cylindrical sandwich shells with metallic foam cores under blast loading-numerical simulations," *Composite Structures*, vol. 99, pp. 213–223, 2013.
- [19] L. Peroni, M. Avalle, and M. Peroni, "The mechanical behaviour of aluminium foam structures in different loading conditions," *International Journal of Impact Engineering*, vol. 35, no. 7, pp. 644–658, 2008.
- [20] W. Nian, K. V. L. Subramaniam, and Y. Andreopoulos, "Dynamic compaction of foam under blast loading considering fluid-structure interaction effects," *International Journal of Impact Engineering*, vol. 50, pp. 29–39, 2012.
- [21] X. Liu, X. Tian, T. J. Lu, D. Zhou, and B. Liang, "Blast resistance of sandwich-walled hollow cylinders with graded metallic foam cores," *Composite Structures*, vol. 94, no. 8, pp. 2485–2493, 2012.
- [22] A. Mortensen and S. Suresh, "Functionally graded metals and metal-ceramic composites: part 1 processing," *International Materials Reviews*, vol. 40, no. 6, p. 239, 1995.
- [23] A. Pollien, Y. Conde, L. Pambaguian, and A. Mortensen, "Graded open-cell aluminium foam core sandwich beams," *Materials Science and Engineering: A*, vol. 404, no. 1-2, p. 9, 2005.
- [24] Y. Conde, A. Pollien, and A. Mortensen, "Functional grading of metal foam cores for yield-limited lightweight sandwich beams," *Scripta Materialia*, vol. 54, no. 4, p. 539, 2006.
- [25] A. H. Brothers and D. C. Dunand, "Mechanical properties of a density-graded replicated aluminum foam," *Materials Science & Engineering A*, vol. 489, no. 1-2, pp. 439–443, 2008.
- [26] J. L. Yu, J. R. Li, and S. S. Hu, "Strain-rate effect and microstructural optimization of cellular metals," *Mechanics of Materials*, vol. 38, no. 1-2, pp. 160–170, 2006.
- [27] S. Li, X. Li, Z. Wang, G. Wu, G. Lu, and L. Zhao, "Finite element analysis of sandwich panels with stepwise graded aluminum honeycomb cores under blast loading," *Composites Part A: Applied Science and Manufacturing*, vol. 80, pp. 1–12, 2016.
- [28] H. Zhou, Y. Wang, X. Wang, Z. Zhao, and G. Ma, "Energy absorption of graded foam subjected to blast: a theoretical approach," *Materials & Design*, vol. 84, pp. 351–358, 2015.
- [29] R. Dou, S. Qiu, Y. Ju, and Y. Hu, "Simulation of compression behavior and strain-rate effect for aluminum foam sandwich panels," *Computational Materials Science*, vol. 112, pp. 205–209, 2016.
- [30] G. W. Ma, Z. Q. Ye, and Z. S. Shao, "Modeling loading rate effect on crushing stress of metallic cellular materials," *International Journal of Impact Engineering*, vol. 36, no. 6, pp. 775–782, 2009.
- [31] K. Li, X.-L. Gao, and G. Subhash, "Effects of cell shape and strut cross-sectional area variations on the elastic properties of three-dimensional open-cell foams," *Journal of the Mechanics and Physics of Solids*, vol. 54, no. 4, pp. 783–806, 2006.
- [32] Z. Zheng, C. Wang, J. Yu, S. R. Reid, and J. J. Harrigan, "Dynamic stress-strain states for metal foams using a 3D cellular model," *Journal of the Mechanics and Physics of Solids*, vol. 72, pp. 93–114, 2014.
- [33] D. Weaire and J. P. Kemrode, "On the distribution of cell areas in a Voronoi network," *Philosophical Magazine B*, vol. 53, pp. 101–105, 1986.
- [34] X. Liu, X. Tian, T. Lu, D. Zhou, and B. Liang, "Blast resistance of sandwich-walled hollow cylinders with graded metallic foam cores," *Composite Structures*, vol. 94, pp. 2489–2493, 2012.
- [35] D. S. Cheng, C. W. Hung, and S. J. Pi, "Numerical simulation of near-field explosion," *Tamkang Journal of Science and Technology*, vol. 16, no. 1, pp. 61–67, 2013.
- [36] J. Kang, S. Shi, and C. Jin, "Theoretical analysis of attenuating shock wave pressure of aluminum foam," *Vibration and Shock*, vol. 29, no. 12, pp. 128–131, 2010.

## Research Article

# Dynamic Response of a Semiactive Suspension System with Hysteretic Nonlinear Energy Sink Based on Random Excitation by means of Computer Simulation

Hui Chen  and Wuyin Jin 

*School of Mechanical & Electromechanical Engineering, Lanzhou University of Technology, Lanzhou 730050, China*

Correspondence should be addressed to Wuyin Jin; [wuyinjin@hotmail.com](mailto:wuyinjin@hotmail.com)

Received 12 July 2020; Revised 18 August 2020; Accepted 26 August 2020; Published 17 September 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Hui Chen and Wuyin Jin. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper aims to investigate the property and behavior of the hysteretic nonlinear energy sink (HNES) coupled to a half vehicle system which is a nine-degree-of-freedom, nonlinear, and semiactive suspension system in order to improve the ride comfort and increase the stability in shock mitigation by using the computer simulation method. The HNES model is a semiactive suspension device, which comprises the famous Bouc–Wen (B-W) model employed to describe the force produced by both the purely hysteretic spring and linear elastic spring of potentially negative stiffness connected in parallel, for the half vehicle system. Nine nonlinear motion equations of the half vehicle system are derived in terms of the seven displacements and the two dimensionless hysteretic variables, which are integrated numerically by employing the direct time integration method for studying both the variables of vertical displacements, velocities, accelerations, chassis pitch angle, and the ride comfort and driver safety, respectively, based on the bump and random road inputs of the pseudoexcitation method as excitation signal. Simulation results show that, compared with the HNES model and the magnetorheological (MR) model coupled to the half vehicle system, the ride comfort and stability have been evidently improved. A successful validation process has been performed, which indicated that both the ride comfort and driver safety properties of the HNES model coupled to half vehicle significantly improved.

## 1. Introduction

To achieve both ride comfort and driver safety is the main objective of the vehicle industry, and there are two ways to achieve ride comfort of high quality; to this end, the first way is to minimize the axis and angular acceleration of the gravity center of the vehicle body, while the second way is to maintain the tire contacts with the ground while the tire strikes a bump. Minimizing the vertical displacement of the vehicle can facilitate the ride stability.

Passive, active, and semiactive suspension vehicle systems are three available classifications of suspensions which depend on the ability of the suspension system to absorb, add, or extract energy. More recently, one of the oldest passive vibration control devices is the tuned mass damper (TMD) which was proposed by Watts in 1883 [1] and patented by Frahm [2], consisting of a mass, a spring, and a viscous

damper which is attached to a primary vibrating system in order to suppress undesirable vibrations. The property of the TMD can be significantly reduced for the environmental influences, and other external parameters may alter its properties and disturb its tuning [3]. To help reduce the vibration, a large oscillating mass is generally required. Both construction and placement are rather difficult. In light of the above drawbacks, to overcome the disadvantages of the TMD, new nonlinear strategies in vibration absorption have been introduced, and among them, researchers investigated the targeted energy transfer (TET) through different designs of NES, so they have gained more attention and are more prevalent [4]. According to [5], definitely, the NES is a nonlinear attachment coupled to a linear system which dissipates energy irreversibly through nonlinear stiffness elements. To help dissipate energy effectively, compared with the TMD, the NES does not need to be tuned to a special

frequency. The performance of nonsmooth NES in the presence of gravity has also been investigated where the pure cubic nonlinearity is broken in vertical configurations as the gravity induces asymmetry to the system [6, 7].

The concept of hysteretic nonlinear energy sink (HNES) was proposed by Tsiatas and studied by Tsiatas and Charalampakis in 2018 [8]. The HNES is the modification of the NES [9, 10], i.e., apart from a small mass and a nonlinear elastic spring of the duffing oscillator, the HNES also comprises a purely hysteretic spring and a linear elastic spring of potentially negative stiffness, connected in parallel. The BoucWen (B-W) model is used to describe the force produced by the purely hysteretic and linear elastic springs and has been investigated in NES dynamics previously [11, 12].

Eltantawie has studied the mechanical behavior of the magnetorheological damper (MRD) as a semiactive device of the vehicle suspension system for decentralized neurofuzzy control method previously. In this paper, the HNES model is employed as a semiactive suspension device and coupled to the linear half vehicle system in order to improve the ride comfort and increase the stability.

This study proposes a comprehensive design by using the HNES model as the semiactive suspension device, and the motion equations of the vehicle system which consist of the HNES model have been established, followed by employing the new direct time integration method, which is presented for solution of motion equations describing the dynamic response of multi-degree-of-freedom, structural linear and nonlinear systems in 2014 [13], for the solution of the system of first-order ordinary differential equations which represent semidiscrete diffusion equations in 2016 [14], and for solution of the motion equations of the vehicle model by means of the computer simulation method.

The rest of this paper is organized as follows: Section 2 describes and analyzes both half vehicle system and the HNES coupled to a primary model. The differential equations of the HNES coupled to a primary model and the semiactive suspension system are analyzed by employing the new direct time integral method, which be formulated in Section 3. The responses of the displacements, velocity, acceleration, and chassis pitch angle in the time domain are simulated by the parameters of half vehicle system in Section 4. The ride comfort and drive safety of half vehicle system are solved numerically, which are described in the Section 5. The main conclusions and findings are drawn in Section 6.

## 2. Half Vehicle and the HNES Model

**2.1. Half Vehicle Coupling with the HNES Model.** In this paper, a semiactive suspension system of the half vehicle model, which consists of sprung mass supported on front and rear suspensions as shown in Figure 1, is proposed. Both the front and rear suspensions which are considered as unsprung masses with the front and rear tire mass are connected to their tire axles, respectively. Damper  $C_f$ , elastic device  $K_f$  and HNES model (details of the HNES model is described in Section 2.2) are connected through chassis mass  $M_2$  and front tire directly in parallel, same as the front suspension, damper

$C_r$ , elastic device  $K_r$  and HNES model connected through chassis mass  $M_2$  and rear tire in parallel. The model is a nine-degree-of-freedom system and represented by the displacements of vertical driver  $U_5$ , vertical vehicle body  $U_4$ , body pitch  $U_3$ , front and rear tire deflections  $U_{f1}$  and  $U_{r1}$  and  $U_{f2}$  and  $U_{r2}$  of the mass  $M_1$ , and the displacements  $U_{Zf}$  and  $U_{Zr}$  which indicates the BoucWen model (Figure 1).

The mathematical model is derived by applying Newton's second law; the motion equations can be formulated as follows:

$$M_3\ddot{U}_5 + C_{f2}(\dot{U}_5 - (\dot{U}_4 - \dot{U}_3L_3)) + K_{f2}(U_5 - (U_4 - U_3L_3)) = 0, \quad (1)$$

$$\begin{aligned} M_2\ddot{U}_4 - C_{f2}(U_5 - (\dot{U}_4 - \dot{U}_3L_3)) - K_{f2}(U_5 - (U_4 - U_3L_3)) \\ + K_f(U_4 - U_3L_1 - U_{f1}) + C_f(\dot{U}_4 - \dot{U}_3L_1 - \dot{U}_{f1}) \\ + C_2(\dot{U}_4 - \dot{U}_{f2} - \dot{U}_3L_1) + K_{NL}(U_4 - U_{f2} - U_3L_1)^3 \\ + K_r(U_4 - U_{r1} + U_3L_2) + C_r(\dot{U}_4 - \dot{U}_{r1} + \dot{U}_3L_1) \\ + C_2(\dot{U}_4 - \dot{U}_{r2} + \dot{U}_3L_2) + K_{NL}(U_4 - U_{r2} + U_3L_2) \\ = -F_{BWf} - F_{BWr}, \end{aligned} \quad (2)$$

$$\begin{aligned} M_1\ddot{U}_{f2} - C_2(\dot{U}_4 - \dot{U}_{f2} - \dot{U}_3L_1) - K_{NL}(U_4 - U_{f2} - U_3L_1)^3 \\ + C_1(\dot{U}_{f2} - \dot{U}_{f1}) + K_{EL}(U_{f2} - U_{f1}) = F_{BWf}, \end{aligned} \quad (3)$$

$$\begin{aligned} M_1\ddot{U}_{r2} - C_2(\dot{U}_4 - \dot{U}_{r2} + \dot{U}_3L_2) - K_{NL}(U_4 - U_{r2} + U_3L_2)^3 \\ + C_1(\dot{U}_{r2} - \dot{U}_{r1}) + K_{EL}(U_{r2} - U_{r1}) = F_{BWr}, \end{aligned} \quad (4)$$

$$\begin{aligned} M_f\ddot{U}_{f1} - K_f(U_4 - U_{f1} - U_3L_1) - C_f(\dot{U}_4 - \dot{U}_{f1} - \dot{U}_3L_1) \\ - C_1(\dot{U}_{f2} - \dot{U}_{f2}) - K_{EL}(U_{f2} - U_{f1}) + K_{tr}(U_{f1} - q_1) = 0, \end{aligned} \quad (5)$$

$$\begin{aligned} M_r\ddot{U}_{r1} - K_r(U_4 - U_{r1} + U_3L_2) - C_r(\dot{U}_4 - \dot{U}_{r1} + \dot{U}_3L_2) \\ - C_1(\dot{U}_{r2} - \dot{U}_{r1}) - K_{EL}(U_{r2} - U_{r1}) + K_{tr}(U_{r1} - q_2) = 0, \end{aligned} \quad (6)$$

$$\begin{aligned} J\ddot{U}_3 + C_{f2}(\dot{U}_5 - (\dot{U}_4 - L_3\dot{U}_3))L_3 + K_{f2} \\ \cdot (U_5 - (U_4 - L_3U_3))L_3 - K_f((U_4 - L_1U_3) - U_{f1})L_1 \\ - C_f((\dot{U}_4 - L_1\dot{U}_3) - \dot{U}_{f1})L_1 \\ - C_2((\dot{U}_4 - L_1\dot{U}_3) - \dot{U}_{f2})L_1 - K_{NL} \\ \cdot ((U_4 - L_1U_3) - U_{f2})^3L_1 + K_r((U_4 + L_2U_3) - U_{r1})L_2 \\ + C_r((\dot{U}_4 + L_2\dot{U}_3) - \dot{U}_{r1})L_2 + C_2 \\ \cdot ((\dot{U}_4 + L_2\dot{U}_3) - \dot{U}_{r2})L_2 + K_{NL}((U_4 + L_2U_3) - U_{r2})^3L_2 \\ = -F_{BWf}L_1 - F_{BWr}L_2, \end{aligned} \quad (7)$$

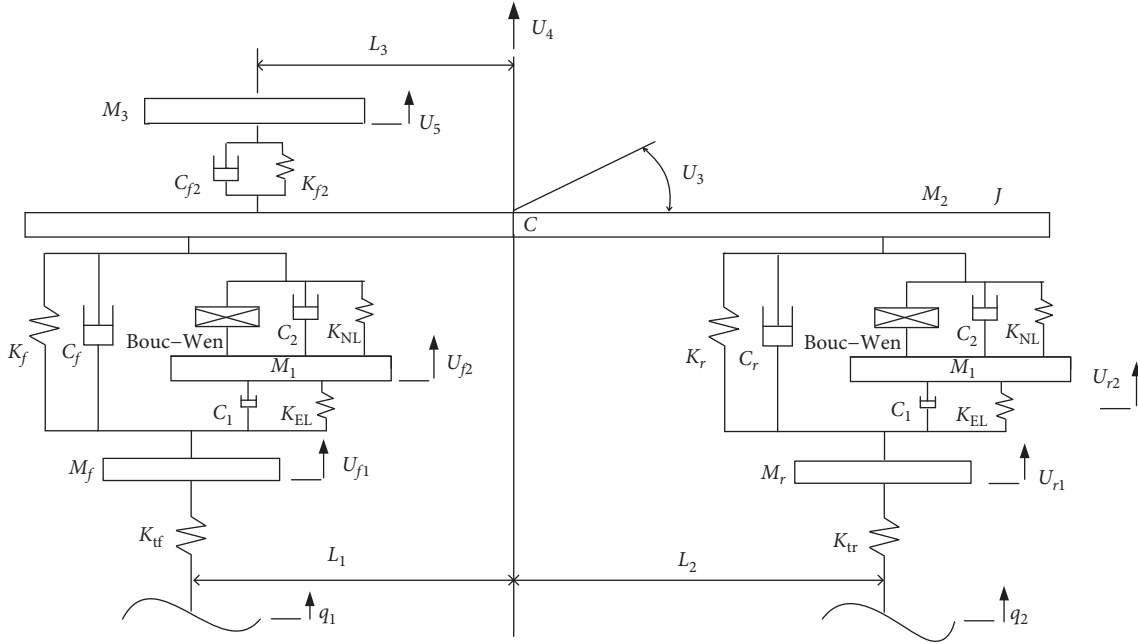


FIGURE 1: Half vehicle system coupling with the HNES model.

$$F_{\text{BWf}} = \alpha\kappa(U_4 - U_3L_1 - U_{f2}) + (1 - \alpha)\kappa DZ_f(t), \quad (8)$$

$$F_{\text{BWr}} = \alpha\kappa(U_4 + U_3L_2 - U_{r2}) + (1 - \alpha)\kappa DZ_r(t), \quad (9)$$

$$\begin{aligned} \dot{Z}_f(t) = & D^{-1}(\dot{U}_4 - \dot{U}_3L_1 - \dot{U}_{f2}) \\ & \cdot \left( (1 - |Z_f(t)|^n)(\gamma \text{sgn}(\dot{U}_4 - \dot{U}_3L_1 - \dot{U}_{f2})Z_f(t) + \beta) \right), \end{aligned} \quad (10)$$

$$\begin{aligned} \dot{Z}_r(t) = & D^{-1}(\dot{U}_4 + \dot{U}_3L_2 - \dot{U}_{r2})(1 - |Z_r(t)|^n) \\ & \cdot (\gamma \text{sgn}(\dot{U}_4 + \dot{U}_3L_2 - \dot{U}_{r2})Z_r(t) + \beta). \end{aligned} \quad (11)$$

Equations (1)–(11) can be rewritten in the matrix form as follows:

$$[\mathbf{M}]\{\ddot{\mathbf{U}}\} + [\mathbf{C}]\{\dot{\mathbf{U}}\} + [\mathbf{K}]\{\mathbf{U}\} = [\mathbf{F}]\{\mathbf{Q}\}, \quad (12)$$

where  $F_{\text{BWf}}$  and  $F_{\text{BWr}}$  are the BoucWen damper forces of the front and rear of the half vehicle system, respectively,  $\mathbf{M}$ ,  $\mathbf{C}$ , and  $\mathbf{K}$  denote matrices of mass, damper, and spring stiffness,  $\ddot{\mathbf{U}}$ ,  $\dot{\mathbf{U}}$ , and  $\mathbf{U}$  are vectors of acceleration, velocity, and displacements, and  $\mathbf{F}$  and  $\mathbf{Q}$  denote excitation vector coefficient matrix and the road inputs vector of the half vehicle system, respectively. For the definition of matrices and vectors  $\mathbf{M}$ ,  $\mathbf{C}$ ,  $\mathbf{K}$ ,  $\ddot{\mathbf{U}}$ ,  $\dot{\mathbf{U}}$ ,  $\mathbf{U}$ ,  $\mathbf{F}$ , and  $\mathbf{Q}$ , refer to Appendix A, and for expressing intuitively in the figures of this paper, the subscript of symbols  $\ddot{U}_n$  and  $\dot{U}_n$  are represented by  $A_n$  and  $V_n$ ,  $n = 1, 2, \dots, 9$ , respectively.

**2.2. Dynamic Responses of Single-Degree-of-Freedom Primary System Coupling with HNES.** The schematic representation of single-degree-of-freedom linear primary system coupling with HNES, which is considered as the HNES model, is

shown in Figure 2. The HNES model is constituted apart from a small mass  $m_2$  and a nonlinear elastic spring of the duffing oscillator; it also comprises a purely hysteretic spring and a linear elastic spring of potentially negative stiffness, which constitute the BoucWen model, connected in parallel. The BoucWen model is used to describe the force produced by the purely hysteretic and linear elastic springs. In light of the above, the single-degree-of-freedom (SDoF) primary system coupling with HNES (the HNES model) is constituted from the BoucWen model, a linear damp  $C_2$ , and a cubic nonlinear spring  $K_{\text{NL}}$  in parallel. Damper  $C_1$  and elastic linear spring  $K_{\text{EL}}$  are connected to mass  $m_1$  in parallel. The BoucWen model, damper  $C_2$ , and cubic nonlinear spring  $K_{\text{NL}}$  are connected between the masses  $m_2$  and  $m_1$  of the constituted attachments of the HNES in parallel.

Assume that the rear tire travels over the same path as the front tire except for a time delay  $\tau$  which can be mentioned as equation (25) in this study. The excitation function of road inputs  $q_1$  and  $q_2$ , which is a stationary random process, acts on the center of the tire and road surface contact point. The BoucWen model, which was introduced by Bouc [15] and extended by Wen [16], is employed to describe the force produced by both the purely hysteretic and linear elastic springs.

The SDoF primary system coupling with HNES (the HNES model) is shown in Figure 2; the hysteresis force is expressed by referring to [17, 18]:

$$F_{\text{BW}} = \alpha kx(t) + (1 - \alpha)k Dz(t), \quad (13)$$

where  $k > 0$ ,  $D > 0$ ,  $t$  is the time history of the input variable, and  $z(t)$  is a dimensionless hysteretic variable which is governed by the following differential equation:

$$\dot{z}(t) = D^{-1}(A - (\beta \text{sgn}(z(t)\dot{x}) + \gamma)|z(t)|^n)\dot{x}, \quad (14)$$

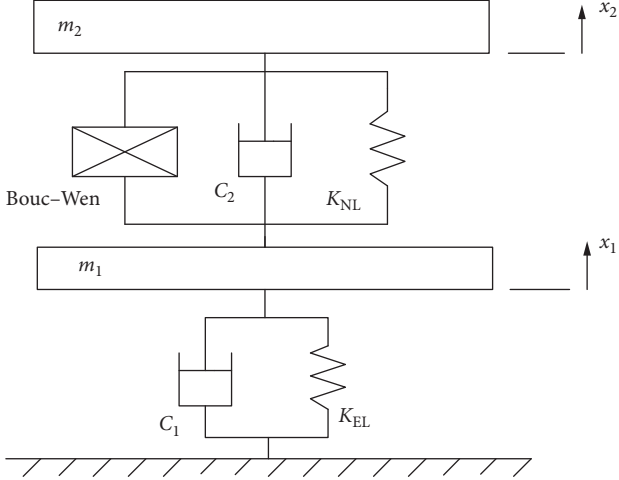


FIGURE 2: The SDoF primary system coupling with HNES (the HNES model).

where  $n > 0$ ,  $\text{sgn}(\cdot)$  denotes the signum function,  $x$  is the displacement,  $k$  is the initial stiffness,  $\alpha$  is the ratio of postyield to preyield stiffness,  $D$  indicates the yield displacement, the dimensionless exponential parameter  $n$  governs the abruptness of transition between preyield and postyield response, and the values of  $\alpha, A, \beta, D, \gamma$ , and  $n$  are set in Table 1 [8].

The HNES model is derived by applying Newton's second law, and the equations of motion can be formulated as follows:

$$m_1 \ddot{x}_1 + C_2 \dot{x}_1 + K_{EL} x_1 - F_{BW} - C_2 (\dot{x}_2 - \dot{x}_1) - K_{NL} (x_2 - x_1)^3 = 0, \quad (15)$$

$$m_2 \ddot{x}_2 + F_{BW} + C_2 (\dot{x}_2 - \dot{x}_1) + K_{NL} (x_2 - x_1)^3 = 0, \quad (16)$$

$$\dot{z}(t) = D^{-1} (\dot{x}_2 - \dot{x}_1) (1 - |z(t)|^n (\text{sgn}(\dot{x}_2 - \dot{x}_1) z(t)) + b). \quad (17)$$

In this investigation, only the direct impulsive force of the primary system is considered, and thus, the initial conditions of the problem are as follows:

$$\begin{aligned} \dot{x}_1(0) &= v_0, \\ x_1(0) = \dot{x}_2(0) = \dot{x}_2(0) &= 0, \end{aligned} \quad (18)$$

where  $m_1, x_1, K_{EL}$ , and  $C_1$  are the mass, displacement, stiffness of spring, and damping coefficient of the linear primary system, respectively, and  $m_2, x_2, K_{NL}$ , and  $C_2$  are the mass, displacement, damping coefficient, and cubic nonlinear stiffness of the HNES's attachments.

### 3. State Space Form of the Motion Equations and the Excitation of Road Inputs

**3.1. State Space Form.** Equations (15)–(18) can be reorganized into state space form as follows:

TABLE 1: Parameters of HNES (dimensionless).

| Parameter | $\alpha$ | $A$ | $\beta$ | $D$ | $\gamma$ | $n$ |
|-----------|----------|-----|---------|-----|----------|-----|
| Value     | -0.3     | 1   | 0.5     | 0.1 | 0.5      | 1   |

$$\{\mathbf{X}_n\} = \begin{cases} X_1 = x_1 \\ X_2 = \dot{x}_1 \\ X_3 = x_2 \\ X_4 = \dot{x}_2 \\ X_5 = z(t) \end{cases}, \quad (19)$$

$$\left\{ \begin{aligned} \dot{X}_1 &= X_2 \\ \dot{X}_2 &= \frac{F_{BW} - C_1 X_2 - K_{EL} X_1 + C_2 (X_4 - X_2) + K_{NL} (X_2 - X_1)^3}{m_1} \\ \dot{X}_3 &= X_4 \\ \dot{X}_4 &= \frac{-F_{BW} - C_2 (X_4 - X_2) - K_{NL} (X_3 - X_1)^3}{m_2} \\ \dot{X}_5 &= D^{-1} (X_4 - X_2) (1 - |X_5|^n (\gamma \text{sgn}(X_4 - X_2) X_2) + \beta) \end{aligned} \right\}. \quad (20)$$

Together with the pertinent initial conditions, the following can be expressed:

$$\{\dot{X}_n(0)\} = \left\{ \begin{aligned} \dot{X}_1(0) &= v_0 \\ \dot{X}_2(0) &= \frac{(C_1 + C_2)v_0}{m_1} \\ \dot{X}_3(0) &= \frac{C_2 v_0}{m_2} \\ \dot{X}_4(0) &= \frac{C_2 v_0}{m_2} \\ \dot{X}_5(0) &= \frac{v_0}{D} \end{aligned} \right\}, \quad (21)$$

where  $v_0 = \dot{x}_1(0)$  denotes the initial value.

Equations (19)–(21) can be rewritten as follows:

$$\begin{aligned} \mathbf{M}_i^{-1} \mathbf{F}(t) - \mathbf{M}_i^{-1} \mathbf{P}(\mathbf{X}_n(t), \dot{\mathbf{X}}_n(t)) \\ = f(\mathbf{X}_n(t) + \mathbf{X}_n(t), t) = \mathbf{0}, \quad n = 1, 2, \dots, 5, \end{aligned} \quad (22)$$

where  $\mathbf{M}_i$ ,  $i = 1, 2$ , denotes the mass matrix which consists of  $m_1$  and  $m_2$  and function  $f(\mathbf{X}_n(t) + \mathbf{X}_n(t), t)$  is already expressed in equation (20).

It should be noteworthy that equation (22) is a typical nonlinear initial value problem, which are solved numerically using the new direct time integration method

introduced by Katsikadelis [13], and that for multidegree of freedom systems is described as follows:

$$\mathbf{X}_{n-k} = \dot{X}_{n-k-1}(t) + \frac{h}{2}\dot{X}_{n-k}(t) + \frac{h}{2}\dot{X}_{n-k-1}(t), \quad (23)$$

where  $\mathbf{X}_{n-k}$  denotes the responses of the displacement vector, of which elements are the value in different time histories for  $k = 1, 2, \dots, (T/h)$ .  $T$  is the simulation period, and  $h$  is the sampling interval.  $\dot{X}_{n-k}$  and  $\mathbf{X}_{n-k}$  can be solved by coupling equations (22) and (23) with  $k = 1, 2, \dots, T/h$ .

Figure 3 depicts the responses of displacements  $X_1$  and  $X_3$  of the masses  $m_1$  and  $m_2$  of primary system coupling with HNES against the impulse magnitude  $v_0 = 0.35$  m/s,  $h = 0.01$ , and  $T = 500$  seconds (Figure 3(a)) and  $v_0 = 2$  m/s,  $h = 0.01$ , and  $T = 500$  seconds (Figure 3(b)). It can be observed that displacements  $X_1$  and  $X_3$  against time are similar in shape during 500 seconds, and the amplitude changes more slowly over the entire time domain, and the maximum values of amplitudes are  $X_1 = 2.1$  m and  $X_2 = 2.3$  m (Figure 3(b)) and  $X_1 = 0.12$  m and  $X_3 = 0.01$  m (Figure 3(a)).

### 3.2. Excitation of Road Inputs

**3.2.1. Bump Road Inputs.** The bump road excitation method is one of the simple inputs for a vehicle, which has the equations as follows [19]:

$$q_i(t) = \begin{cases} \frac{u_r(1 - \cos(8\pi t))}{2}, & 0.5 \leq t \leq 0.75, \\ 0, & \text{otherwise,} \end{cases} \quad (24)$$

where  $i (=1, 2)$  represents the front and rear tire, respectively,  $u_r = 0.11$  m denotes the amplitude of bump road excitation, simulation period  $T = 4$  seconds, the bump road excitation for the rear tire is assumed to be that of the front but with a time delay (seconds) of  $\tau$ , and sampling interval  $h = 0.01$  seconds in this study. Time delay  $\tau$  is formulated as follows:

$$\tau = \frac{L_1 + L_2}{v}, \quad (25)$$

where symbol  $v$  denotes the vehicle velocity and  $L_1$  and  $L_2$  are the front and rear wheelbase; the excitation signals for bump road inputs are shown in Figure 4(a).

The road excitation of rear tire inputs  $q_2(t)$  is formulated as

$$q_2(t) = q_1(t - \tau). \quad (26)$$

**3.2.2. Random Excitation of Road Surface Roughness.** The road power spectral density (PSD) about the spatial frequency is  $G_q(n)$ ,  $n_1 < n < n_2$  ( $n = f/v$ ), where  $v$  is the average vehicle velocity and  $n$  is the spatial frequency. The road surface roughness variance of  $\sigma_q^2$  by using the frequency spectrum properties of average stationary random process can be expressed as follows [20]:

$$\sigma_q^2 = \int_{n_1}^{n_2} G_q(n) dn. \quad (27)$$

Dividing the area ( $n_1, n_2$ ) into  $m$  smaller intervals, the frequency of the power spectral density value is  $G_q(n_{\text{mid-}i})$  instead of  $G_q(n)$  at each small interval of the central part, where  $n_{\text{mid-}i}$ ,  $i = 1, 2, \dots, m$ . Throughout, the area between the value of the formula after discretization approximation can be written as follows:

$$\sigma_q^2 \approx \sum_{i=1}^m G_q(n_{\text{mid-}i}) \cdot \Delta n_i. \quad (28)$$

Corresponding to each small range, we need to find the frequency for  $n_{\text{mid-}i}$  ( $i = 1, 2, 3, \dots, m$ ) and the standard deviation of  $\sqrt{G_q(n_{\text{mid-}i}) \cdot \Delta n_i}$  sine wave function, and such sine wave function can be expressed as

$$\sqrt{G_q(n_{\text{mid-}i}) \cdot \Delta n_i} \cdot \sin(2\pi n_{\text{mid-}i} x + \theta_i). \quad (29)$$

Then, the corresponding sine wave function in each area is added up to obtain the road random displacement input on the time domain:

$$q_j(t) = \sqrt{2} \sum_{i=1}^m \sqrt{G_q(n_{\text{mid-}i}) \cdot \Delta n_i} \cdot \sin(2\pi n_{\text{mid-}i} t + \theta_i), \quad j = 1, 2, \quad (30)$$

where  $\theta_i$  is uniformly distributed random numbers between  $[-\pi, \pi]$ , of which probability density and distribution function can be referred to in Appendix B. And,  $t$  denotes the time history which can be normalized by  $s/v$  ( $s$  is the total distance traveled by the vehicle in this study). According to reference [21], the spatial frequency  $n$  of road is distributed between  $[0.011, 2.83] \text{ m}^{-1}$ . In this study, considering both the relevant regulations and practical life experience, the range of vehicle velocity  $V_v \in [V_{n-\text{min}}, V_{n-\text{max}}] = [10, 30] \text{ m/s}$  (36–108 km/h).

Relationships between time frequency  $f$  and spatial frequency  $n$  of the vehicle can be formulated as equation  $f = vn$  (Hz). Therefore, the range of road spatial frequency in this paper can be expressed as follows:

$$\begin{aligned} \mathbf{A}_n &= \mathbf{A}_{n1} \cap \mathbf{A}_{n2} = [V_{n-\text{min}} \times 0.011, V_{n-\text{min}} \times 2.83] \\ &\cap [V_{n-\text{max}} \times 0.011, V_{n-\text{max}} \times 2.83] = [0.33, 28.3]. \end{aligned} \quad (31)$$

Here,  $\mathbf{A}_{n1}$  and  $\mathbf{A}_{n2}$  denote the spatial frequency range at lowest velocity  $V_{n-\text{min}}$  and at the highest velocity  $V_{n-\text{max}}$  and symbol  $\cap$  denotes a mathematical operation for the intersection of two sets.

The random excitation signal of road surface roughness is shown in Figure 4(b).

## 4. Simulation Response in the Time Domain

Equations (1)–(11) can be altered into equation (12) which is a nonlinear equation and can be written using the direct time integration method numerically as follows:



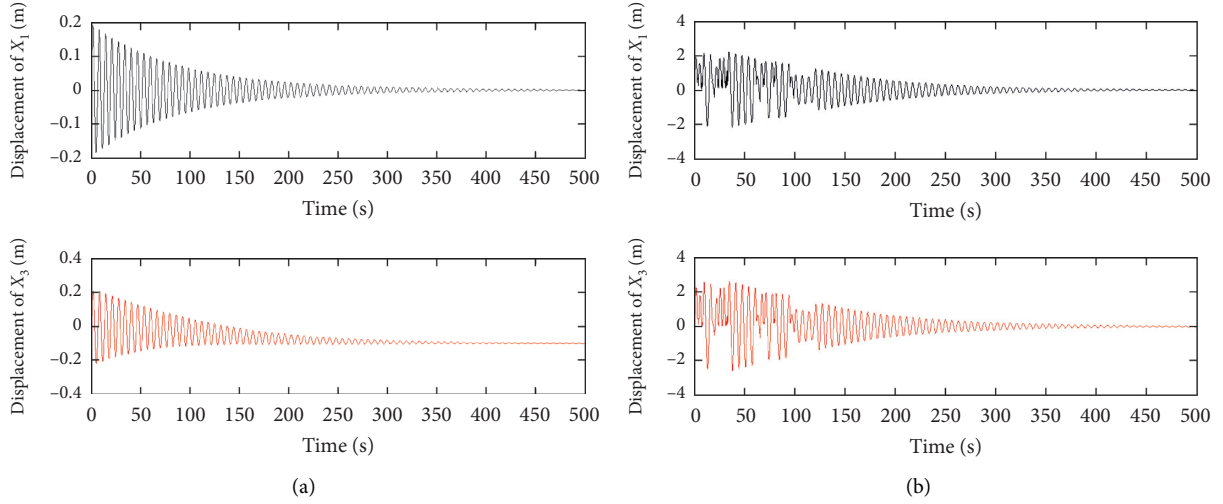


FIGURE 3: The response of displacements of the primary system coupling with the HNES model for  $v_0 = 0.35$  m/s,  $h = 0.01$ , and  $T = 500$  seconds (a) and  $v_0 = 2$  m/s,  $h = 0.01$ , and  $T = 500$  seconds (b).

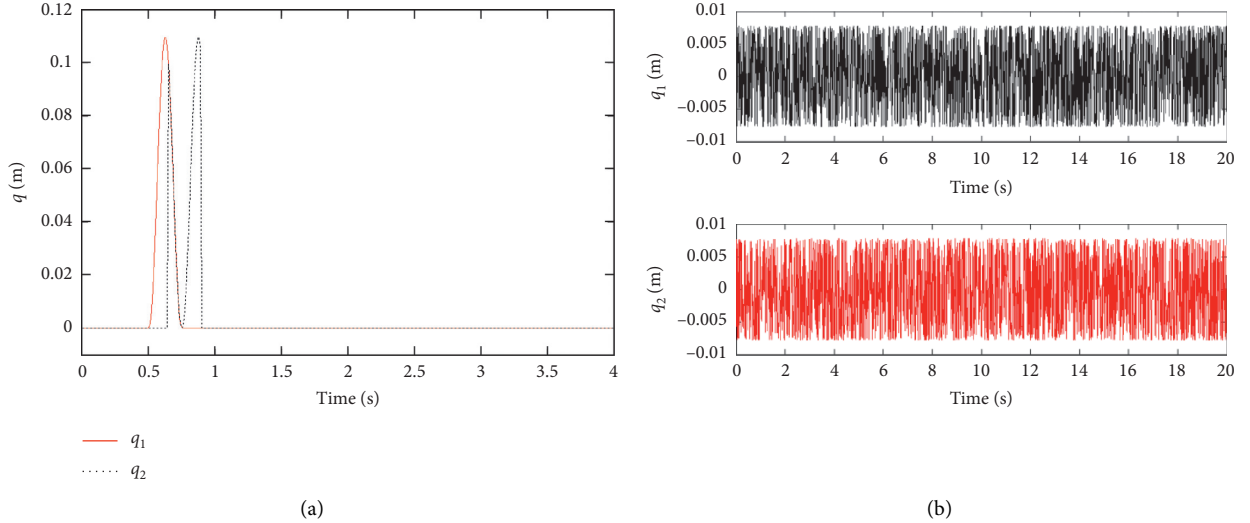


FIGURE 4: Two types of excitation signal for (a) the bump road inputs, where simulated period  $T = 4$  seconds, sampling interval  $h = 0.01$  second, and vehicle velocity  $v = 20$  m/s, and (b) the random road inputs, where simulated period  $T = 20$  seconds, sampling interval  $h = 0.01$  seconds, and vehicle velocity  $v = 20$  m/s.

$$\mathbf{M}\ddot{\mathbf{U}}_n + \mathbf{F}(\dot{\mathbf{U}}_n, \mathbf{U}_n) = \mathbf{P}_n, \quad (32)$$

$$\begin{bmatrix} -h\mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \dot{\mathbf{U}}_n \\ \mathbf{U}_n \end{Bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \dot{\mathbf{U}}_{n-1} \\ \mathbf{U}_{n-1} \end{Bmatrix} + \begin{bmatrix} \frac{\mathbf{r}_1}{2}\mathbf{I} \\ \frac{\mathbf{r}_2}{2}\mathbf{I} \end{bmatrix} \ddot{\mathbf{U}}_n + \begin{bmatrix} \frac{\mathbf{r}_1}{2}\mathbf{I} \\ \frac{\mathbf{r}_2}{2}\mathbf{I} \end{bmatrix} \ddot{\mathbf{U}}_{n-1}, \quad (33)$$

$$\ddot{\mathbf{U}}_0 = \mathbf{M}^{-1}(\mathbf{P}_0 - \mathbf{F}(\dot{\mathbf{U}}_0, \mathbf{U}_0)), \mathbf{U}(0) = \mathbf{U}_0, \dot{\mathbf{U}}(0) = \dot{\mathbf{U}}_0, \quad \det(\mathbf{M}) \neq 0, \quad (34)$$

where  $r_1 = (h_2/2)$ ,  $r_2 = h$ ,  $h$  is the sampling interval time (second), and size  $(\mathbf{I}) = [9 \times 9]$  is the unit matrix.

$\ddot{\mathbf{U}}_n$ ,  $\dot{\mathbf{U}}_n$ , and  $\mathbf{U}_n$  can be solved by combining equations (32) and (33) with  $n = 1, 2, \dots, T/h$ .  $\dot{\mathbf{U}}_n$  and  $\mathbf{U}_n$  can be solved using equation (33) which is a linear equation and then substituting into equation (32) which is a nonlinear equation and can be solved to yield  $\ddot{\mathbf{U}}_n$ . The function “fsovl” of *Matlab 2017a* has been employed to obtain the numerical results of both equations.

## 5. Performance Index and Numerical Simulation

**5.1. Response of Half Vehicle System Coupling with the HNES Model.** By querying the power spectral density coefficient (PSD) of standard road surface excitation  $G_q(n_0)$ [22] and referring a 10-degree-of-freedom (DOF) passenger vehicle model determine six state variables related to the longitudinal, lateral, and heave velocities and roll, pitch, and yaw rates [23]. Equations (32)–(34) can be solved using the software programming *Matlab 2017a* numerically. The parameters of half vehicle system coupling with the HNES model are given in Table 2.

Figure 5 depicts the response of vertical displacements of driver  $U_5$  and chassis  $U_4$  and pitch angle of chassis  $U_3$  against the road inputs for the bump excitation, of which simulation period  $T=10$  seconds, sampling interval  $h=0.01$ , and different velocities of the vehicle  $v=1.1256$  m/s (Figure 5(a)) and  $v=20$  m/s (Figure 5(b)), respectively. From these figures, it can be observed that for the bump road excitation with the amplitude  $A=0.11$  m, the chassis (sprung mass) reached a maximum of 0.05 m (Figure 5(a)) and 0.09 m (Figure 5(b)). Also when the inputs dropped to zero, the chassis (sprung mass) vertical displacement went down for only  $-0.001$  m (Figure 5(a)) and  $-0.01$  m (Figure 5(b)) which is almost zero. The overall amplitude achieved by the chassis is equal to that against the front and rear tire inputs. The response of the angular displacement of chassis had a maximum peak value of 0.021 rad (Figure 5(a)) and 0.04 rad (Figure 5(b)). However, when the rear tire of the vehicle received the input of road excitation, the maximum displacement response of the driver is  $-0.024$  m while the vehicle velocity is 1.1256 m/s, and  $-0.022$  m while the vehicle velocity is 20 m/s. Both values are smaller than the response values of the model in reference [24]. The output data obtained from Figure 5(a) for the maximum displacement of the vehicle body against time is 0.036 m, and the maximum angular displacement of the vehicle pitch angle is 0.02 rad which is smaller than of the model mentioned in reference [24].

The solution data obtained from the programs of the motion equations of the half vehicle system for velocity of the drive  $V_5$ , chassis (sprung mass)  $V_4$ , and angular velocity of the chassis  $V_3$  are plotted against time in Figure 6. It can be grasped that for bump excitation road inputs, the peak value is 0.11 m, simulation period  $T=10$  seconds, sampling interval  $h=0.01$  seconds, and different velocities of the vehicle  $v=1.1256$  m/s (Figure 6(a)) and  $v=20$  m/s (Figure 6(b)), respectively. When the front tire of the vehicle passed road excitation  $q^1(t)$ , the maximum velocity responses of the vehicle body are 0.21 m/s (a) and 0.272 m/s (b), respectively. When the rear tire excitation  $q^2(t)$  received by the vehicle gradually approaches 0 m, the velocity of the vehicle body only decreases by  $-0.08$  m/s (a) and  $-0.18$  m/s (b). These two values gradually tend to be stable after 10 seconds. The response of the driver velocity  $V_5$  stabilizes faster than others.

The responses of the displacements  $U_5, U_4$ , and  $U_3$ , velocities  $V_5, V_4$ , and  $V_3$ , and the accelerations  $A_5, A_4$ , and  $A_3$  against the road inputs were recorded for both the front and rear tire random excitation road inputs in Figures 7–9. It is obvious that the amplitude of displacement, velocity, and acceleration against time for the vehicle

TABLE 2: Main parameters of half vehicle system coupling with the HNES model.

| Parameter                | Meaning of the symbols   | Value and units        |
|--------------------------|--|------------------------|
| $M_3$                    | Driver mass  | 65 kg                  |
| $M_2$                    | Chassis mass   | 708 kg                 |
| $M_f, M_r$               | Front/rear unsprung mass                                       | 80 kg                  |
| $J$                      | Chassis inertia  | 2700 kg·m <sup>2</sup> |
| $L_1$                    | Horizontal distance from chassis mass center to the front axle | 1.011 m                |
| $L_2$                    | Horizontal distance from chassis mass center to the rear axle  | 1.803 m                |
| $L_3$                    | Horizontal distance from chassis mass center to the driver     | 0.779 m                |
| $U_5$                    | Driver vertical displacement                                   | m                      |
| $U_4$                    | Chassis vertical displacement                                  | m                      |
| $U_3$                    | Chassis pitching angle   | rad                    |
| $U_f, U_r$               | Vertical displacement of the front and rear axis               | m                      |
| $K_{j2}$                 | Stiffness of connecting the chassis to the driver              | 23071 N/m              |
| $K_f, K_r$               | Stiffness of connecting chassis to the front and rear axle     | 1600 N/m               |
| $K_{f\beta}, K_{r\beta}$ | Stiffness of the front and rear tire                           | 16000 N/m              |
| $C_{t2}$                 | Damping coefficient connecting the chassis and driver          | 1500 N·s/m             |
| $C_f, C_r$               | Damping coefficient connecting chassis to front and rear axle  | 980 N·s/m              |
| $C_1$                    | Damping coefficient of the linear primary system               | 0.001 N·s/m            |
| $C_2$                    | Damping coefficient of the HNES attachments                    | 0.01 N·s/m             |
| $m_1$                    | Mass of the linear primary system                              | 1 kg                   |
| $F_{BWf}, F_{BW_r}$      | Force of the hysteretic restoring of the front/rear tire axle  | N                      |
| $K_{NL}$                 | Cubic nonlinear stiffness of the HNES attachment               | 1 N/m <sup>3</sup>     |
| $K_{EL}$                 | Stiffness of the linear primary system                         | 1 N/m                  |
| $U_{j2}, U_{r2}$         | Front and rear tire displacements of the linear primary system | m                      |

velocity in Figure 7(a), Figure 8(a), and Figure 9(a) is smaller than that in Figure 7(b), Figure 8(b), and Figure 9(b), respectively. It can be seen that the displacement of vehicle body had a series random value, of which the maximum is 0.51 m, against the time for road inputs of random excitation, while the velocity of the vehicle is 20 m/s. It can be observed that the response of the displacement, velocity, and the acceleration of half vehicle system coupling with the HNES model against the road inputs for random excitation is very consistent with the vehicle theory, while the velocity is in the range [1.1256, 20] m/s. Therefore, from the dynamic response analysis, it can be known that it is reasonable and reliable to apply the HNES model to the vehicle system.

**5.2. Performance of the Half Vehicle System Coupling with the HNES Model.** Another simulation program has created with the aid of the semiactive suspension of half vehicle system coupling with the HNES model to obtain the drive safety and the ride comfort property. The drive safety and the quality of the contact between the tire and the road are described by

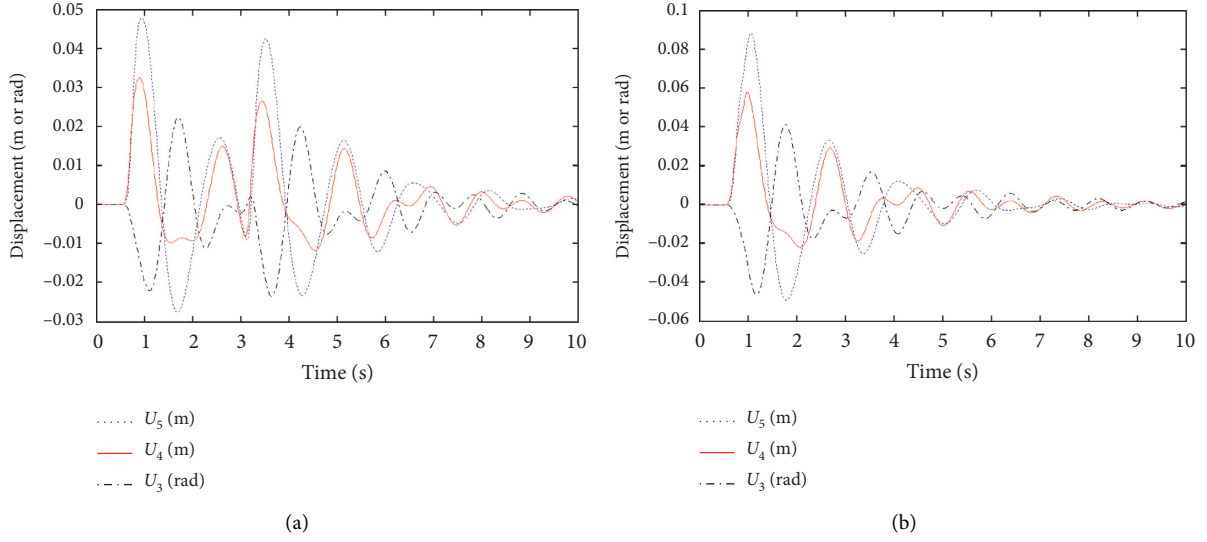


FIGURE 5: Responses of displacement  $\mathbf{U}_n$  of the half vehicle system while the excitation is bump road inputs. The vector  $\mathbf{U}_n$  represents the displacements of the driver  $U_5$ , the chassis  $U_4$ , and the pitch angle  $U_3$ , i.e.,  $\mathbf{U}_n = (U_5, U_4, U_3)^T$ , respectively. Simulation period  $T=10$  seconds, sampling interval  $h=0.01$  seconds, vehicle velocity  $v=1.1256$  m/s (the time delay  $\tau=2.5$  s) and (b) same as (a) except for the vehicle velocity  $v=20$  m/s (the time delay  $\tau=0.0938$  s).

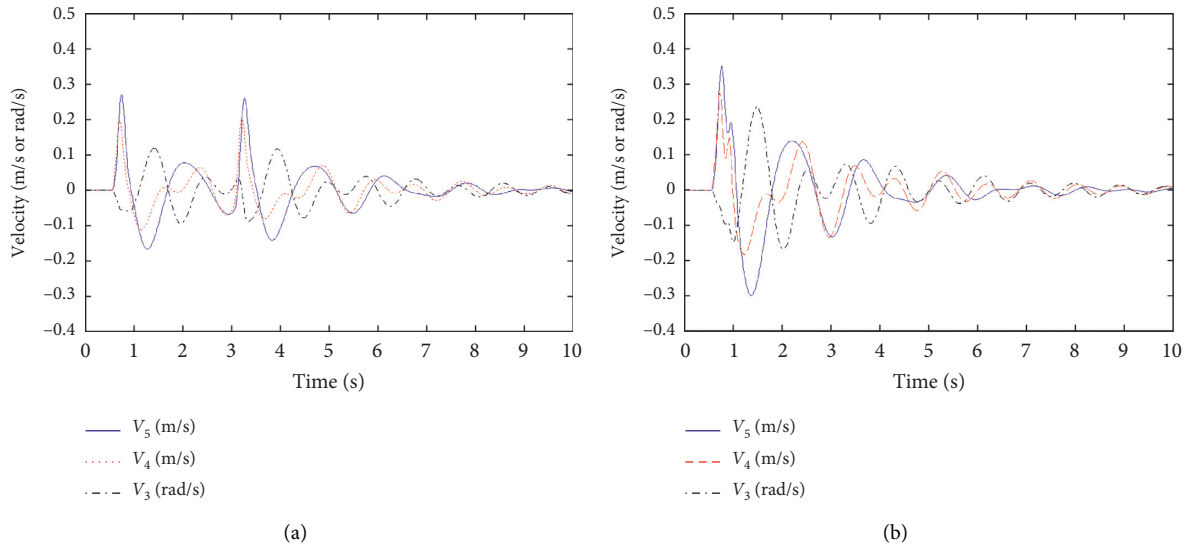


FIGURE 6: Responses of velocities  $\dot{\mathbf{U}}_n$  of half vehicle system coupling with the HNES model while excitation is bump road inputs and the velocity of vehicle is 1.1256 m/s (a) and 20 m/s (b). The vector  $\dot{\mathbf{U}}_n$  presents the velocities of the driver, the chassis, and the pitch angle.

the RMS (root mean square) value of the forces, which are dynamically exchanged by the tire and the road that is defined in [21]:

$$F_{U_{i1}Dyn} = K_{ti}(U_{i1} - q_j). \quad (35)$$

Ride comfort can be formulated as

$$\ddot{U}_4^{RMS} = \sqrt{\frac{1}{T} \int_0^T \left( \frac{\ddot{U}}{g} \right)^2 dt}. \quad (36)$$

Drive safety can be formulated as

$$F_{U_{i1},dyn}^{RMS} = \sqrt{\frac{1}{T} \int_0^T \left( \frac{F_{U_{i1},dyn}}{F_{U_{i1},stat}} \right)^2 dt}. \quad (37)$$

The steady-state load on the tire can be formulated as

$$F_{U_{i1},stat} = (M_1 + M_i + M_2)g, \quad (38)$$

where  $i$  represents the front and rear suspensions, respectively, and  $j=1, 2$ .

Figure 10 shows the ratio of dynamic loads to static loads for the front (rear) tire, and the square of the ratio of

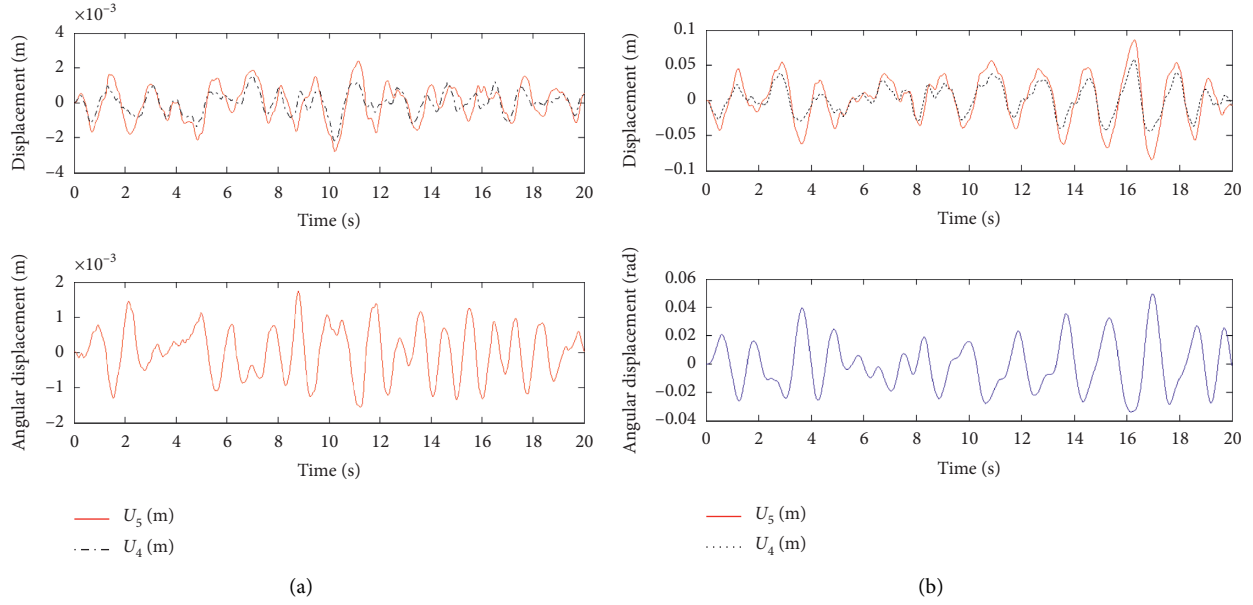


FIGURE 7: Responses of displacements  $U_n$  of half vehicle system coupling with the HNES model while the excitation is random road inputs and the velocity of vehicle is 1.1256 m/s (a) and 20 m/s (b).

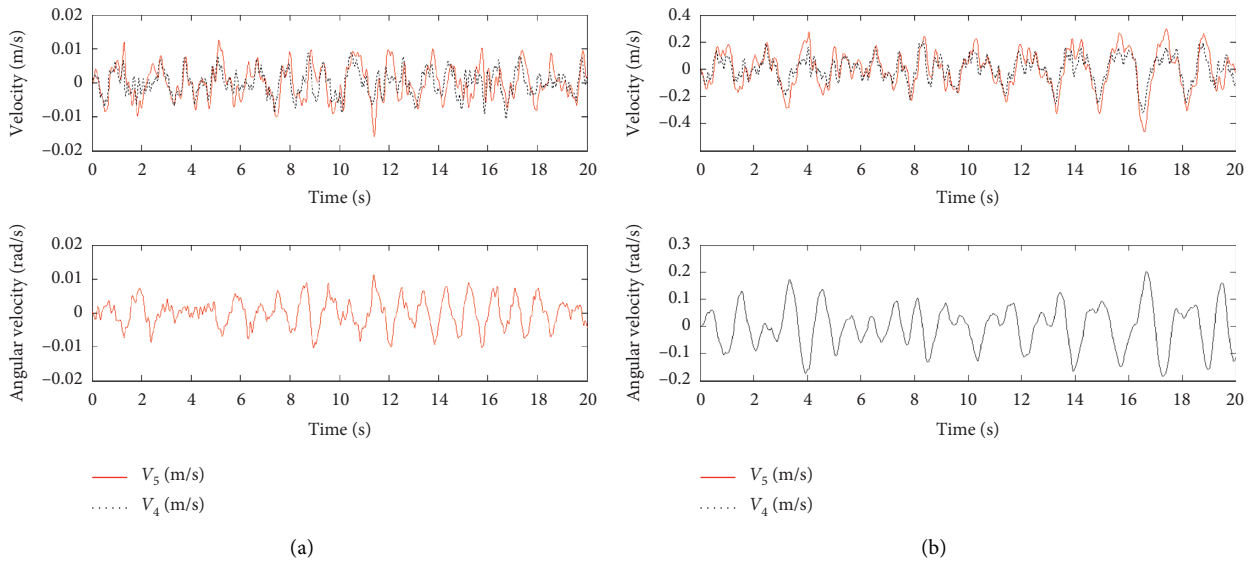


FIGURE 8: Responses of velocities  $V_n$  of half vehicle system coupling with the HNES model while the excitation is random road inputs and the velocity of vehicle is 1.1256 m/s (a) and 20 m/s (b).

acceleration to gravity acceleration  $g$  against the road inputs is recorded for bump road excitation in Figure 10(a) and road inputs for random excitation in Figure 10(b) while the vehicle velocity  $v = 1.1256$  m/s (the delay time is 2.5 seconds). It is obvious that the ratio of dynamic loads to static loads of the front tire had a maximum peak value of 0.48 against the road inputs for bump excitation in the time range of [0.5, 0.75] seconds, of which rear tire had a maximum peak value of 0.058 after the time delay of 2.5 seconds in Figure 10(a). The square of ratio of acceleration to gravity acceleration against the road inputs for bump excitation had a maximum peak value of 0.041 in the time range of [0.7, 0.75] seconds.

The maximum peak value of 0.58 was reached after the delay time of 2.5 seconds in Figure 10(a). The ratio of dynamic loads to static loads of the front (rear) tire, and the square of the ratio of acceleration to gravity acceleration against the road inputs for random excitation is very small which tends to zero as shown in Figure 10(b).

It can be observed from Tables 3 and 4 that the responses of displacements, velocities, and accelerations of the half vehicle system coupling with the HNES model are smaller. The ride comfort and driver safety properties are lower than those in the semiactive suspension of the MR model with the same parameters. However, ride comfort

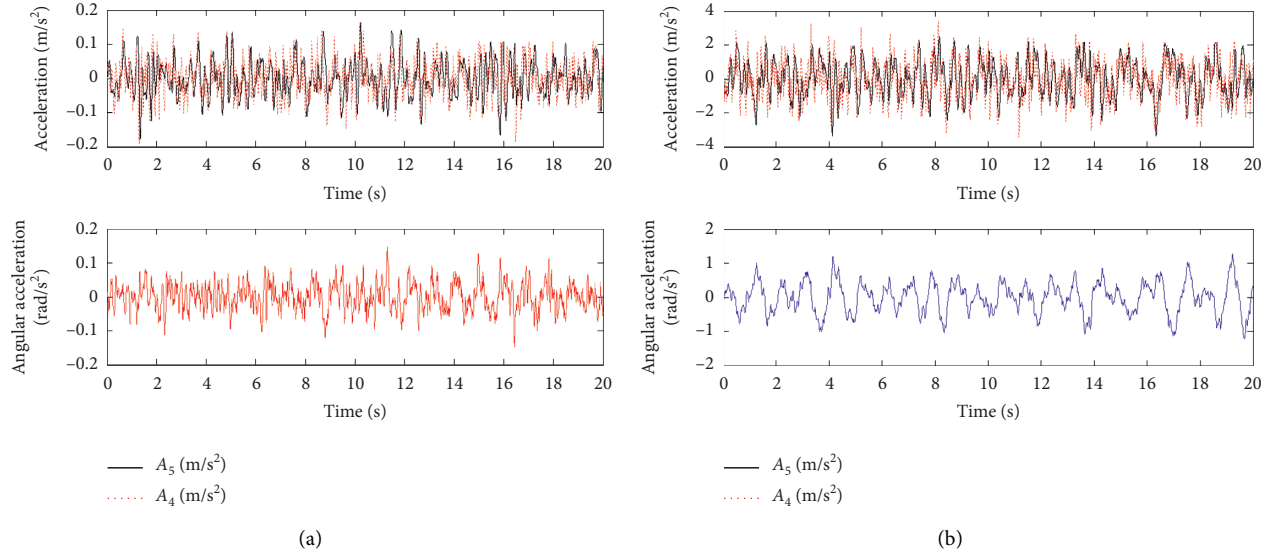


FIGURE 9: Responses of accelerations  $A_n$  of half vehicle system coupling with the HNES model while the excitation is random road inputs and the velocity of vehicle is 1.1256 m/s (a) and 20 m/s (b).

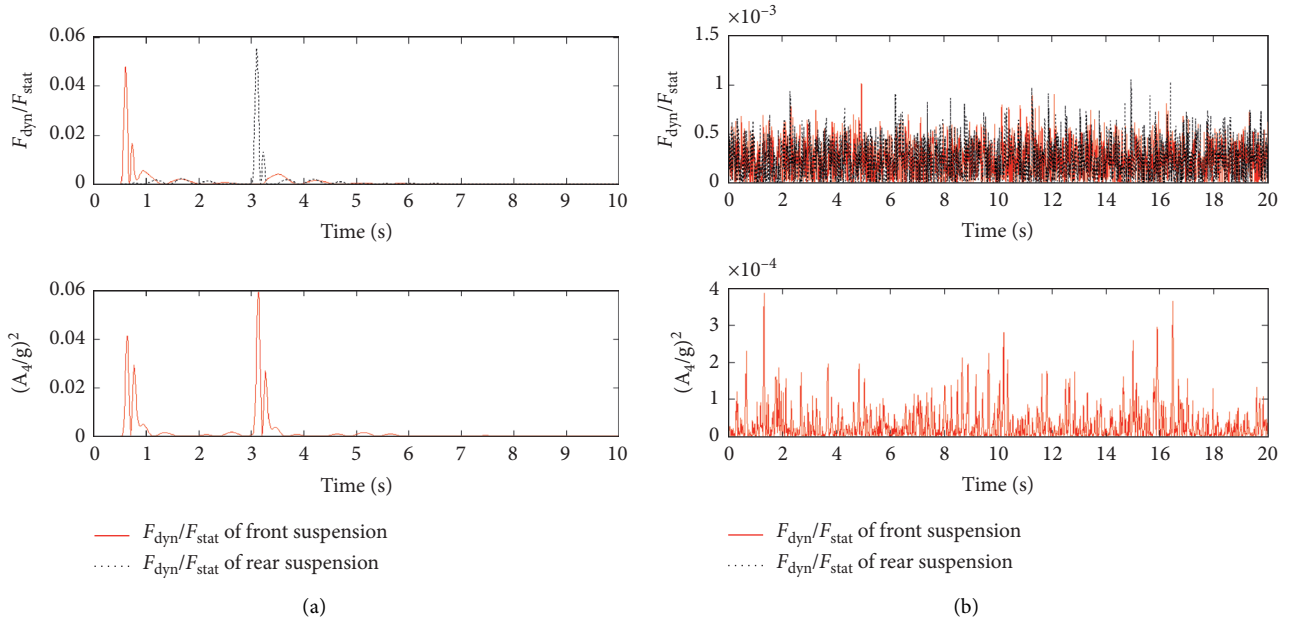


FIGURE 10: Responses of the ratio of dynamic loads and static loads of the front and rear tire against road inputs and the square of ratio of acceleration and gravity acceleration against the road inputs (a) for bump excitation and (b) for random inputs. The vehicle velocity  $v = 1.1256$  m/s (the time delay  $\tau = 2.5$  second), the simulation period  $T = 20$  seconds, and sampling interval ( $h$ ) = 0.01 seconds.

TABLE 3: Comparison of simulation response.

| Items | Half vehicle system coupling with HNES model (this paper) |                            | The model in the reference [23] |                          | Vehicle velocity and time delay |                |
|-------|---|----------------------------|---------------------------------|--------------------------|---------------------------------|----------------|
|       | Bump road inputs  | Random road inputs         | Bump road inputs                | Random road inputs       | $v$                             | $\tau$         |
| $U_4$ | 0.032 m   | $2.1e^{-3}$ m              | 0.04 m                          | 0.038 m                  | 1.1256 m/s                      | $\tau = 2.5$ s |
| $U_3$ | 0.022 rad   | $1.78e^{-3}$ rad           | 0.03 rad                        | 0.04 rad                 |                                 |                |
| $A_4$ | —   | $0.152$ m/s <sup>2</sup>   | —                               | $4.2$ m/s <sup>2</sup>   |                                 |                |
| $A_3$ | —   | $0.089$ rad/s <sup>2</sup> | —                               | $0.5$ rad/s <sup>2</sup> |                                 |                |

TABLE 4: Comparison of both ride comfort and drive safety.

| Items   | The model in the reference [23] |                 | Half vehicle system coupling with the HNES model |           |                 |           |
|---------|---------------------------------|-----------------|--|-----------|-----------------|-----------|
|         | Front suspension                | Rear suspension | Front suspension                                 |           | Rear suspension |           |
| Ride    | Bump inputs                     | 2.28            |  |           | 3.7516e-5       |           |
| Comfort | Random inputs                   | 4.6             |  |           | 2.0739e-4       |           |
| Drive   | Bump inputs                     | 1.34            | 1.96   | 5.3061e-5 |                 | 2.5438e-5 |
| Safety  | Random inputs                   | 5.32            | 4.31   | 3.8679e-4 |                 | 2.5438e-5 |

parameters of the half vehicle system coupling with the HNES model have only been achieved for bump road inputs.

## 6. Conclusions

In this paper, the behavior and performance of a semi-active suspension consisting of the HNES for the half vehicle system has been proposed and investigated in shock mitigation. A successful validation process has been performed to solve the half vehicle system coupling with HNES using the same parameters to compare the developed model to the MR model. The response of the half vehicle system coupling with HNES is solved numerically using the direct time integration method (analog equation method). The model is simulated against two kinds of road inputs which are bump road inputs and random road inputs. The great performance of the model is evident as the velocity of the vehicle changes from 1.1256 m/s to 20 m/s. The main conclusions that can be drawn from this investigation are as follows:

- (i) The HNES model as a semiactive suspension coupled to the linear half vehicle system was studied

- (ii) The performance of semiactive suspension in half vehicle system coupling with the HNES model has a significant potential of reducing the road input disturbance
- (iii) The responses of displacements, velocities, and accelerations of the HENS model as a semiactive suspension in the half vehicle system are much better than those of the MR model as a semiactive suspension system for the same parameter in the time domain ( $v = 1.1256(\text{m/s})$ )
- (iv) The ride comfort and the drive safety of the HNES as a semiactive suspension in the half vehicle system both against the bump road inputs and random road inputs are much higher than those of the MR model given in the study [23]

## Appendix

### A. Matrices M, K, and C and Vectors F, Q, and U in equation (12)

Matrices M, K, and C and vectors F, Q, and U are

$$\begin{aligned}
 \mathbf{M} &= \text{diag}(M_3 \ M_2 \ J \ M_1 \ M_1 \ M_f \ M_r), \\
 \mathbf{K} &= \begin{pmatrix} K_{f2} & -K_{f2} & K_{f2}L_3 & 0 & 0 & 0 & 0 \\ -K_{f2} & K_{f2} + K_{f2} + K_{NL} + K_r + K_{NL} & -K_{f2}L_3 - K_fL_1 - K_{NL}L_1 - K_rL_2 + K_{NL}L_2 & -K_{NL} & -K_f & -K_{NL} & -K_r \\ K_{f2}L_3 & K_{f2}L_3 + K_fL_1 + K_{NL}L_1 + K_rL_2 + K_{NL}L_2 & K_{f2}L_3^2 - K_fL_1^2 - K_{NL}L_1^2 + K_rL_2^2 + K_{NL}L_2^2 & K_{NL}L_1 & -K_fL_1 & -K_{NL}L_2 & -K_rL_2 \\ 0 & -K_{NL} & K_{NL}L_1 & K_{NL} + K_{EL} & -K_{EL} & 0 & 0 \\ 0 & -K_{NL} & -K_{NL}L_2 & K_{NL} + K_{EL} & -K_{EL} & K_{NL} + K_{EL} & -K_{EL} \\ 0 & -K_f & K_fL_1 & -K_{EL} & K_f + K_{EL} + K_{tf} & 0 & 0 \\ 0 & -K_r & -K_rL_2 & 0 & 0 & -K_{EL} & K_e + K_{EL} + K_{tr} \end{pmatrix}, \\
 \mathbf{C} &= \begin{pmatrix} C_{f2} & -C_{f2} & C_{f2}L_3 & 0 & 0 & 0 & 0 \\ -C_{f2} & C_r + C_{f2} + C_f + 2C_2 & -C_{f2}L_3 - C_fL_1 - C_2L_1 + C_rL_2 + C_2L_2 & -C_2 & -C_f & -C_2 & -C_r \\ -C_{f2}L_3 & C_{f2}L_3 + C_fL_1 + C_2L_1 + C_rL_2 + C_2L_2 & -C_{f2}L_3^2 - C_fL_1^2 - C_2L_1^2 + C_rL_2^2 + C_2L_2^2 & -C_2L_1 & -C_fL_1 & C_2L_2 & -C_rL_2 \\ 0 & -C_2 & C_2L_1 & C_2 + C_1 & -C_1 & 0 & 0 \\ 0 & -C_2 & -C_2L_2 & 0 & 0 & C_2 + C_1 & -C_1 \\ 0 & -C_f & C_fL_1 & -C_1 & C_f + C_1 & 0 & 0 \\ 0 & -C_r & -C_rL_2 & 0 & 0 & -C_1 & C_r + C_1 \end{pmatrix},
 \end{aligned}$$

$$\mathbf{F} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ -F_{\text{BWf}} & -F_{\text{BWr}} & 0 & 0 \\ -F_{\text{BWf}} & -F_{\text{BWr}} & 0 & 0 \\ F_{\text{BWf}} & 0 & 0 & 0 \\ 0 & F_{\text{BWr}} & 0 & 0 \\ 0 & 0 & K_{\text{tf}} & 0 \\ 0 & 0 & 0 & K_{\text{tr}} \end{pmatrix}, \quad (\text{A.1})$$

$$\mathbf{Q} = (1 \ 1 \ q_1 \ q_2)^T,$$

$$\mathbf{U} = (U_5 \ U_4 \ U_3 \ U_{f2} \ U_{f1} \ U_{r2} \ U_{r1} \ U_{Zf} \ U_{Zr})^T.$$

## B. Probability Density and Distribution Function of Random Variable $\theta$

The probability density of random variable  $\theta$  is

$$f(\theta) = \begin{cases} \frac{1}{\pi - (-\pi)}, & -\pi < \theta < \pi, \\ 0, & \text{otherwise.} \end{cases} \quad (\text{B.1})$$

The distribution function of random variable  $\theta$  is

$$F(\theta) = \begin{cases} 0, & \theta < -\pi, \\ \frac{\theta - (-\pi)}{\pi - (-\pi)}, & -\pi \leq \theta < \pi, \\ 1, & \theta \geq \pi. \end{cases} \quad (\text{B.2})$$

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the Key Research and Development Program of Gansu Province, China (Grant No. 18YF1GA063).

## References

- [1] P. Watts, "On a method of reducing the rolling of ships at sea. Read at the 24th Session of the Royal Institution of Naval Architects," Report no. 1883-12, RINA Transactions, Genova, Italy, 1883.
- [2] H. Frahm, *Device for Damping Vibrations of Bodies*, Google Patents, Dublin, Ireland, 1911.
- [3] B. Weber and G. Feltrin, "Assessment of long-term behavior of tuned mass dampers by system identification," *Engineering Structures*, vol. 32, no. 11, pp. 3670–3682, 2010.
- [4] M. A. AL-Shudeifat, "Highly efficient nonlinear energy sink," *Nonlinear Dynamics*, vol. 76, no. 4, pp. 1905–1920, 2014.
- [5] E. Boroson, S. Missoum, P.-O. Mattei, and C. Vergez, "Optimization under uncertainty of parallel nonlinear energy sinks," *Journal of Sound and Vibration*, vol. 394, pp. 451–464, 2017.
- [6] A. T. Savadkoohi, C.-H. Lamarque, and Z. Dimitrijevic, "Vibratory energy exchange between a linear and a non-smooth system in the presence of the gravity," *Nonlinear Dynamics*, vol. 70, no. 2, pp. 1473–1483, 2012.
- [7] M. Weiss, C. Mahmoud, S. Alireza Ture, and C. H. Lamarque, "Multi-scale energy exchanges between an elasto-plastic oscillator and a light nonsmooth system with external pre-stress," *Nonlinear Dynamics*, vol. 83, no. 1-2, pp. 109–135, 2016.
- [8] G. C. Tsiatas and A. E. Charalampakis, "A new hysteretic nonlinear energy sink (HNES)," *Communications in Nonlinear Science and Numerical Simulation*, vol. 60, pp. 1–11, 2018.
- [9] A. F. Vakakis and O. Gendelman, "Energy pumping in nonlinear mechanical oscillators: Part II-resonance capture," *Journal of Applied Mechanics*, vol. 68, no. 1, pp. 42–48, 2001.
- [10] O. Gendelman, L. I. Manevitch, A. F. Vakakis, and R. M'Closkey, "Energy pumping in nonlinear mechanical oscillators: Part I-dynamics of the underlying Hamiltonian systems," *Journal of Applied Mechanics*, vol. 68, no. 1, pp. 34–41, 2001.
- [11] A. T. Savadkoohi, C.-H. Lamarque, and M. V. Contessa, "Trapping vibratory energy of main linear structures by coupling light systems with geometrical and material nonlinearities," *International Journal of Non-linear Mechanics*, vol. 80, pp. 3–13, 2016.
- [12] D. Laxalde, F. Thouverez, and J.-J. Sinou, "Dynamics of a linear oscillator connected to a small strongly non-linear hysteretic absorber," *International Journal of Non-linear Mechanics*, vol. 41, no. 8, pp. 969–978, 2006.
- [13] J. T. Katsikadelis, "A new direct time integration method for the equations of motion in structural dynamics," *ZAMM - Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, vol. 94, no. 9, pp. 757–774, 2014.
- [14] J. T. Katsikadelis, "A new direct time integration method for the semi-discrete parabolic equations," *Engineering Analysis with Boundary Elements*, vol. 73, pp. 181–190, 2016.
- [15] R. Bouc, "Forced vibrations of mechanical systems with hysteresis," in *Proceedings of the Fourth Conference on Nonlinear Oscillations*, Prague, Czech Republic, September 1967.

- [16] Y.-K. Wen, "Method for random vibration of hysteretic systems," *Journal of the Engineering Mechanics Division*, vol. 102, no. 2, pp. 249–263, 1976.
- [17] A. Charalampakis and V. Koumousis, "On the response and dissipated energy of Bouc–Wen hysteretic model," *Journal of Sound and Vibration*, vol. 309, no. 3-5, pp. 887–895, 2008.
- [18] A. E. Charalampakis, "The response and dissipated energy of Bouc–Wen hysteretic model revisited," *Archive of Applied Mechanics*, vol. 85, no. 9-10, pp. 1209–1223, 2015.
- [19] G. Smijith and K. Jayaraj, "MR damper and its application in vehicle suspension system," *Mechatronics*, vol. 12, 2010.
- [20] J. Lin and D. Sun, "Application of pseudo excitation method to 3-D buffeting analysis of the Tsing Ma longspan suspension bridge," *Journal-Dalian University of Technology*, vol. 39, pp. 172–179, 1999.
- [21] International Organization for Standardization, *Mechanical Vibration and Shock–Evaluation of Human Exposure to Whole-Body Vibration–Part 5: Method for Evaluation of Vibration Containing Multiple Shocks*, International Organization for Standardization, Geneva, Switzerland, 2004.
- [22] L. Jiahao, "A fast CQC algorithm of PSD matrices for random seismic responses," *Computers & Structures*, vol. 44, no. 3, pp. 683–687, 1992.
- [23] M. Čorić, "Optimisation of active suspension control inputs for improved performance of active safety systems," *Vehicle System Dynamics*, vol. 56, no. 1, pp. 1–26, 2018.
- [24] M. A. Eltantawie, "Decentralized neuro-fuzzy control for half car with semi-active suspension system," *International Journal of Automotive Technology*, vol. 13, no. 3, pp. 423–431, 2012.



## Research Article

# A Deep Learning Approach for a Source Code Detection Model Using Self-Attention

Yao Meng  and Long Liu

State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou 450001, China

Correspondence should be addressed to Yao Meng; mengyao@ncwu.edu.cn

Received 17 July 2020; Revised 24 August 2020; Accepted 1 September 2020; Published 16 September 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Yao Meng and Long Liu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of deep learning, many approaches based on neural networks are proposed for code clone. In this paper, we propose a novel source code detection model At-biLSTM based on a bidirectional LSTM network with a self-attention layer. At-biLSTM is composed of a representation model and a discriminative model. The representation model firstly transforms the source code into an abstract syntactic tree and splits it into a sequence of statement trees; then, it encodes each of the statement trees with a deep-first traversal algorithm. Finally, the representation model encodes the sequence of statement vectors via a bidirectional LSTM network, which is a classical deep learning framework, with a self-attention layer and outputs a vector representing the given source code. The discriminative model identifies the code clone depending on the vectors generated by the presentation model. Our proposed model retains both the syntactics and semantics of the source code in the process of encoding, and the self-attention algorithm makes the classifier concentrate on the effect of key statements and improves the classification performance. The contrast experiments on the benchmarks OJClone and BigCloneBench indicate that At-LSTM is effective and outperforms the state-of-art approaches in source code clone detection.

## 1. Introduction

In modern society, the application of computers and software has already pervasively permeated our lives. The necessities of life such as medical care, resources, communication, and public security depend on the running of the software of high quality. However, the development of software has always been very costly without adequate knowledge of our world. Software engineers are always busy fixing defects and refactoring code during the life cycle of software. The evolution of software tools and development process has never ceased. The industry has been exploring new methods to reduce the software complexity, improve the development efficiency, and enhance the reliability and maintainability of software.

Programmers often involve similar code with common operations such as copy-paste action, using framework, and generating code by automatic tools. The moderate use of code clone plays an active role in the software life cycle,

aiming at reducing errors and speeding software development. According to the research, code clone is common in software systems. There is around 15% to 25% similar code in Linux kernel [1], while in published JDK packages, people find more than 29% similar code [2].

However, the abuse of code clone is a huge challenge in the process of software testing, maintaining, and redevelopment, which may spoil code readability, duplicate code defects, and even bring in malicious code without intention [3, 4]. Especially, in the phase of software maintaining, ubiquitous clone of poor quality becomes the nightmare of programmers. The modification of duplicate code might expand the size of code base. The study on the code clone detection is hence attracted by scholars in software engineering. People proposed detection algorithms based on text, tokens, code structure, and program graphs successively.

Code clone, also known as similar code or duplicate code, refers to two or more identical or similar code snippets

in the source code library. The famous scholar Bellon [5] divides clone pairs into four categories in his paper. Type 1 and Type 2 are almost identical code segments, while Type 2 and Type 3 are heterogeneous code pairs with similar structure or semantics. The corresponding definitions are given in Section 3.1.

The traditional clone detection algorithms are mostly implemented with natural language processing technology. These approaches usually treat the source code as a plain text or a sequence of tokens and then calculate the similarity by means of line by line mapping or word vectors' distance [2, 6]. In addition, scholars [7, 8] try to represent the source code with other techniques such as the latent semantic index [9] and hash mapping [10], in order to improve the detection efficiency of the model. However, these representation approaches are based on NLP methods, which mean only simple clone pairs, i.e., type 1 and type 2 clones, are detectable.

In recent years, with the rapid development of deep learning technology, researchers in the field of code cloning detection begin to analyze the structural and semantic features of source code with deep learning approaches. Based on the common neural networks such as convolutional neural network, recursive neural network, and recurrent neural network, scholars exploit various deep encoders [11–13] to automatically extract deep features from the Abstract Syntactical Trees (ASTs), which are generated from the original source code. Compared to the traditional methods, deep neural networks are capable of extracting the features hidden in the AST, which improves the classification performance in detecting semantic cloning.

The LSTM network, which was invented in 1997, is a classical recurrent neural network framework. It contains a memory cell which can preserve states over long periods of time. The structure of LSTM is detailed in Section 3. Due to the ability of handling sequential data, more and more scholars tend to use LSTM to locate cloning code in software engineering. However, the training of the deep encoder based on LSTM is still affected by long-term dependency, as the hierarchy of the whole AST is too deep.

In this paper, we propose a source code clone detection model At-BiLSTM using bidirectional LSTM with self-attention mechanism, which retains both the syntactics and semantics of code in the process of representation. At-BiLSTM is a static analytical model based on the deep learning framework, including a source code representation model and a discriminative model. The representation model firstly transforms the source code into an AST and then splits it into a sequence of statement trees according to a preorder traversal algorithm. Each divided tree represents a legitimate programming statement in the source code, corresponding to a subtree rooted by a statement node in the whole AST. By the decomposition of AST, our model greatly reduces the risk of long-term dependence existing in other AST analytic models [12, 13] based on recurrent neural networks. The representation model then converts the statement trees into an ordered set of vectors via a specific LSTM encoder. Finally, the representation model obtains a single vector representation as an output from these

statement vectors based on a bidirectional LSTM encoder with a self-attention layer. Depending on self-attention mechanism, the output vector indicates a weighted average of all statements in the source code, in which core statements are enhanced with heavy weights. A discriminative model is attached to the output of the representation model in At-BiLSTM. It gives a reasonable prediction of code clone by comparing the code similarity in a supervised learning way.

As far as we know, we are the first to propose a clone detection model based on self-attention mechanism. The main contributions of our work are as follows:

- (i) We propose a novel neural source code representation model, which maps the AST, transferred from the source code, to a high dimensional space. Compared to the previous research, the output vector in our model retains the source code structure and semantics.
- (ii) Based on the decomposition of the whole AST, our model effectively reduces the depth and complexity of the original tree and alleviates the negative impact on the classification accuracy due to the long-term dependence and the gradient vanishing.
- (iii) To the best of our knowledge, we are the first to introduce self-attention mechanism, which was developed in NLP, into the area of code cloning. With the self-attention mechanism, the core statements in the source code are strengthened, which finally improve the classification performance.
- (iv) The achievements are made when we apply representation model in the detection of code clone. We improve the model parameters with supervised learning on dataset benchmarks. The experiments show that our model is superior to the state-of-art approaches.

The rest of this paper is organized as follows. Section 2 presents the related work in this area. Section 3 describes the notations and background in code clone. Section 4 describes the detailed design of our approach. Section 5 presents the comparison experiments; in Section 6, we make an analysis on the experiment results. Section 7 presents the threats to validity. Finally, we conclude the paper in Section 7.

## 2. Related Work

*2.1. Traditional Approaches.* By the analysis of text and tokens, most of the traditional code detection approaches treat source code as plain text or set of tokens, ignoring its structure and semantics. Such models often extract features manually or with NLP methods in the process of vectorization and then identify clone pairs via text similarity algorithm. Mostly applied in the early stage, these models were capable of identifying clone pairs of type 1 and type 2. Due to the low complexity and cross-platform features, these approaches are widely used in industry.

In the early studies, Mayrand et al. [14] defined specific sample features from the function names, expressions, and

control flow in the source code by manual extraction. They calculated the similarity of clone code pairs by comparing 21 function-level metrics. Based on the features generated manually, the approach strictly relied on the domain knowledge of experts. The adaptability of the model is poor as the structural features of various programming languages are different.

A famous plug-in in Eclipse platform, SDD [15], which is based on token detection, locates clone snippets in source code. Employing the inverted index and N-nearest neighbor algorithms, it effectively reduces the detection time, helping developers quickly locate the clone code in the IDE environment. SDD makes great achievements in the detection of type 1 and type 2 code, while it fails to identify clone pairs which are highly dissimilar syntactically but still perform the same function.

Another famous model Deckard [16], which is widely used for contrast experiments by later researchers, made an analysis on the AST, which is generated from the given the source code. Deckard manually generates the vectors with predefined rules when traversing the tree. After processing the vectors with local sensitive hash algorithm (LSH), the model calculates the similarity of the code according to Euclidean distance and finally outputs the probability of the clone pairs. Deckard might detect clone code pairs with statements in different orders because the corresponding AST retains the basic structure of source code, and it can be applied in different programming platforms with corresponding AST parsers. However, Deckard fails to be implemented in large-scale clone detection due to the complexity of clustering operations.

*2.2. Deep Learning Approaches.* With the development of deep learning theory, people tend to solve problems in every field with deep learning approaches. In recent years, researchers in the field of software engineering have introduced deep neural networks such as multilayer perceptron network, recurrent neural network, and convolutional neural network to solve the problem of code clone detection. Models with deep neural networks can automatically extract the structural and semantic features hidden in the source code. And the classifiers in these models, which are carefully trained by supervised or unsupervised learning, might have the ability to identify the clone pairs of type 3 or type 4.

White et al. [13] firstly employed recursive autoencoder to learn latent features for code clone. They transferred each leaf node in the AST into a feature vector via a word-embedding algorithm and then obtained the vectors of the nonterminal nodes with a bottom-up traversal algorithm recursively. The vector of the root became the representation of the whole AST at the end of the tree traversal. Finally, they identified the clone pairs by comparing the similarity of root vectors. In order to reduce calculating cost, the AST was transformed to the complete binary tree for encoding, which might cause the tree depth even heavier. Their model performed well in a small, author-defined dataset.

The unsupervised learning model often has worse classification performance than the supervised learning

model. Inspired by White's model, Wei et al. [12] developed a supervised learning model called CDLH. They made contrast experiments by sampling data from the famous code clone dataset BigCloneBench [17]. Their basic idea was to exploit AST-based LSTM to extract code features automatically, meanwhile use the similarity in functional behaviors as the supervised information to guide the deep feature learning process. Different from White's model, they vectorized the leaf nodes of AST by word2vector [18], which was very common in NLP. In order to reduce learning time, they calculated the similarity of hamming space instead of vector pairs, which was generated from binary hash functions, to identify the code clone. The experimental results showed their model was effective and well surpassed the previous models based on unsupervised learning.

Besides the detection models based on LSTM networks, researchers also attempted to obtain the semantic information from source code with special CNN. Yu et al. [11] generated both the original AST and the enhanced AST based on token information from the same code snippet, and then they extracted the feature vectors from these two trees via a specific convolution network. They innovatively used the triangle convolution kernel with a depth of 2 to traverse the corresponding trees layer by layer. In order to reduce the UNK values when encoding the AST leaf nodes, they proposed a new algorithm called PACE that combined one-hot encoding with the position of characters for vectorization.

Although recent work provides strong evidence that AST-based detection models can obtain better classification performance, they have two problems. First, these tree-based approaches are vulnerable to gradient vanishing and long-term dependency, especially when the size of grammar trees is very big and deep. Second, most approaches based on AST parsing either transform AST to or directly treat AST as a full binary tree for simplification or efficiency, which destroys the original structure of source code and makes the tree deeper. In order to solve these problems, Zhang et al. [19] proposed an ASTNN model in 2019. They split the AST into a sequence of small statement trees and encoded them with a recursive encoder. Finally they produced a vector representation of the given AST with a bidirectional GRU encoder [20] based on the sequence of statement vectors. They applied the representation approaches to the task of source code classification and code clone detection. Their experiments showed that ASTNN was superior to previous approaches.

The deep learning detection models based on AST parsing employ both the lexical and semantical information hidden in the source code, improving classifiers' performance. Moreover, the extraction of these features is highly automatic by the deep representation layers, which reduces the cost of expert intervention. The detection models determine the granularity of clone detection by the pruning of corresponding AST, which is not trivial with traditional approaches. Therefore, the AST-based traversal algorithms with deep learning have attracted more and more attention in the field of source code clone.

In addition to the detection approaches above, researchers have also exploited graph-based detection [21, 22]

models to detect clone pairs of type 3 and type 4. They used program dependency graphs (PDGs), which contained data flow graph and program control graph, to detect code clone in a semidynamic way. These approaches eventually transformed the problem of clone detection into the problem of finding isomorphic subgraphs over PDGs. Compared to the AST parsing detection, graph parsing detection technology relies on PDG generators, which cannot be used in various platforms easily. These models, often involved with graph mapping or hybrid mapping, are extremely difficult, which researchers fail to deploy in large-scale code base due to the computing cost.

### 3. Preliminaries

In this section, we briefly give some well-accepted definitions in the area of code clone and deep learning.

**3.1. Code Clone.** We give some definitions of code clone as follows:

- (i) Code fragments: also called code snippets, a continuous segment of source code, specified by the triple  $(l, s, e)$ , including the source file  $l$ , the line the fragment starts on,  $s$ , and the line it ends on,  $e$ .
- (ii) Clone pair: a pair of code fragments that are similar, specified by the triple  $(f_1, f_2, \emptyset)$ , including the similar code fragments  $f_1$  and  $f_2$ , and their clone type  $\emptyset$ .
- (iii) Clone class: a set of code fragments that is similar. Specified by the tuple  $(f_1, f_2, \dots, f_n, \emptyset)$ , each pair of distinct fragments is a clone pair:  $(f_i, f_j, \emptyset)$ ,  $i, j \in 1..n$ ,  $i \neq j$ .
- (iv) Earlier studies did not provide a clear classification of code clones until Bellon grouped code clones into 4 types, whose classification has been widely used in later studies. In this paper, we also employ their definitions of clone types.
- (v) Type 1 (T1): syntactically identical code fragments, except for differences in white space and comments.
- (vi) Type 2 (T2): in addition to Type 1 clone differences, syntactically identical code fragments, except for differences in identifier names and literal values.
- (vii) Type 3 (T3): in addition to Type 2 clone differences, syntactically similar code fragments that differ at the statement level. These fragments can have statements added, modified, or removed with respect to each other.
- (viii) Type 4 (T4): syntactically dissimilar code fragments that are still the same semantically. Bellon did not give an accurate definition of T4 clones. It is recognized in the industry that two code fragments with similar functionality, even if they are completely different in structure, are recognized as T4 clones. For example, one code fragment

implementing bubble sort and another code fragment implementing quick sort are considered a pair of T4 clones.

**3.2. Abstract Syntax Tree.** Abstract Syntax Tree is a kind of syntax tree representing the abstract syntactic structure of the source code. It has been widely used by programming compilers and software engineering tools due to the powerful representation ability. Different from concrete syntax trees (CTS), abstract syntax trees do not contain all the details of source code such as the punctuation and delimiters. They only include the syntax structure of the source code at an abstract level. AST contains both the lexical and syntax information of the source code, which is often employed in the industry as an intermediate tool to extract the hidden information.

Figure 1 depicts the structure of an AST. The left part of the graph is a simple IF statement written in Java, while the right part is an AST generated from the IF statement. Nodes of the AST are corresponding to the constructs or symbols of the source code. We conclude from the graph that the AST fully retains the structural information of the source code. However, the depths of the ASTs are usually very deep according to the grammar rules. Recent studies of code clone detection are mainly based on the traversal algorithms of the AST [19, 23, 24].

**3.3. Attention.** The attention mechanism is the internal process for machines to imitate the human observation of things in the world. When processing images, our vision quickly obtains the target area by a global scan, i.e., the focus of attention. Then, the brain pays more attention to the focus area for details, while suppressing irrelevant information.

Bahdanau et al. [25] firstly applied the attention mechanism in the field of machine translation, i.e., sequence to sequence learning. They successfully solved the problem of long-term dependency in machine translation, i.e., the translation models based on fixed vector representations often lost history information of long sentences in the decoding process. After that, people begin to study deep encoders with attention layers, which achieved amazing effects in all areas of NLP.

Mathematically, the attention is essentially a mapping function composed of Query, Key, and Value. The calculation of attention values is divided into three steps:

Step 1: Query is combined with each Key to calculate the attention weight. The similarity function  $f(Q, K)$  can be defined in multiple ways. The simplest calculation is shown as

$$f(Q, K) = K^T Q. \quad (1)$$

Step 2: the softmax function is used to normalize the attention weight obtained in step 1, as shown in equation (2). Sometimes, we have to scale the attention scores later as the original ones are too large to calculate:

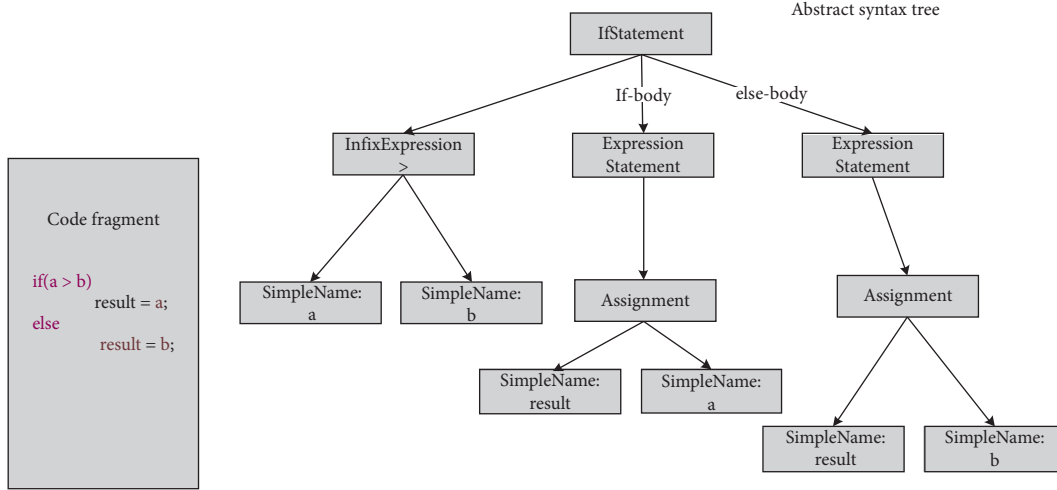


FIGURE 1: Abstract syntax tree.

$$a_i = \text{softmax}(f(Q, K)). \quad (2)$$

Step 3: the final attention is the weighted sum of normalized weights  $a_i$  and corresponding values, which is shown as

$$\text{Attention}(Q, K, V) = \sum a_i V. \quad (3)$$

The attention mechanism is widely used in RNN-based encoder-decoder models. The input of the decoder's current state is determined by the weighted average of all hidden layers' output values in the encoder. The attention algorithm is transformed to the self-attention algorithm when  $Q=K=V$  in the encoder. Our representation model in At-biLSTM is implemented with a specific self-attention encoder layer.

**3.4. LSTM Network.** Recurrent neural networks (RNN) are able to process input sequences of arbitrary length via the recurrent structure with shared weights. Unfortunately, a common problem with the traditional RNN is that components of the gradient vector can grow or decay exponentially over long sequences during training. Therefore, the LSTM architecture, which was invented by Hochreiter and Schmidhuber in 1997 [26], addresses this problem of learning long-term dependency by introducing a memory cell that is able to preserve states over long periods of time. The core part of the LSTM is a cell memory  $C_t$ , including an input gate  $i_t$ , an output gate  $o_t$ , and a forget gate  $f_t$ . Figure 2 describes the basic structure of LSTM. Different from the original RNN, the LSTM solves the problem of long-term dependency effectively with the special structure memory cell, discarding the trivial information in the history to avoid the gradient vanishing [27, 28].

The cell memory of LSTM is composed of three gate controllers. The forget gate  $f_t$  controls the extent to which the previous memory cell is forgotten, which is essential a sigmoid function. The entry of  $f_t$  is a weighted combination

of the previous output value and the current input value. The input gate  $i_t$  controls how much each unit is updated, and the output gate  $o_t$  controls the exposure of the internal memory state. The LSTM transition equations are the following:

$$\begin{aligned} f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \\ C_t &= f_t \times C_{t-1} + i_t \times \tanh(W_f \cdot [h_{t-1}, x_t] + b_c), \\ o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \\ h_t &= o_t \cdot \tanh(C_t), \end{aligned} \quad (4)$$

where  $W_i$ ,  $W_f$ , and  $W_o$  are the weighted matrices and  $b_i$ ,  $b_f$ , and  $b_o$  are the biases of LSTM to be learned during training, parameterizing the transformations of the input, forget, and output gates, respectively.  $\sigma$  and  $\tanh$  are the activation functions, and  $\cdot$  denotes the element-wise multiplication.  $x_t$  is the input of the LSTM cell unit, and  $h_t$  is the output of the hidden layer at the current time step.

Since LSTM has an amazing effect in dealing with long sequential data, researchers propose several variants, e.g., Gated Recurrent Unit (GRU), Bidirectional LSTM (Bi-LSTM), and Bidirectional GRU (Bi-GRU). These networks usually improve the performance when handling the sequential data.

## 4. Our Proposed Approach

In this section, we present our model At-biLSTM, which is an end-to-end learning approach that unifies the representation model and discriminative model. Figure 3 summarizes the overall architecture of At-biLSTM.

**4.1. General Framework.** As shown in Figure 3, At-biLSTM mainly contains two parts: the representation model and the discriminative model. As the core part of the detection system, the representation model transforms the source code into high-dimensional vectors in three steps.

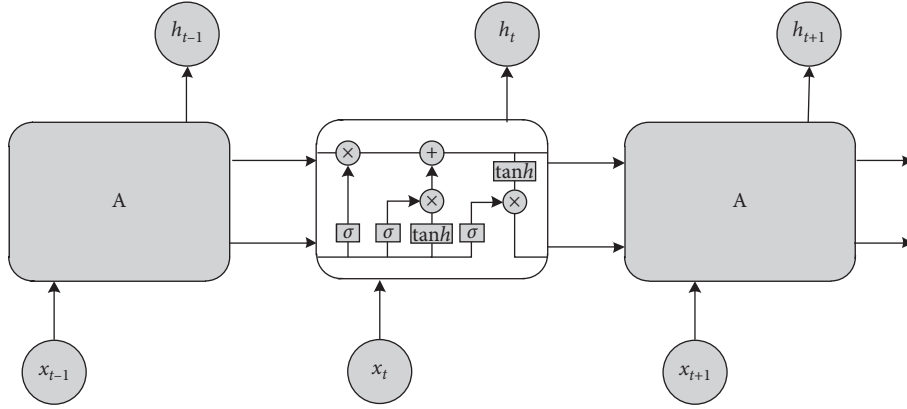


FIGURE 2: Structure of LSTM.

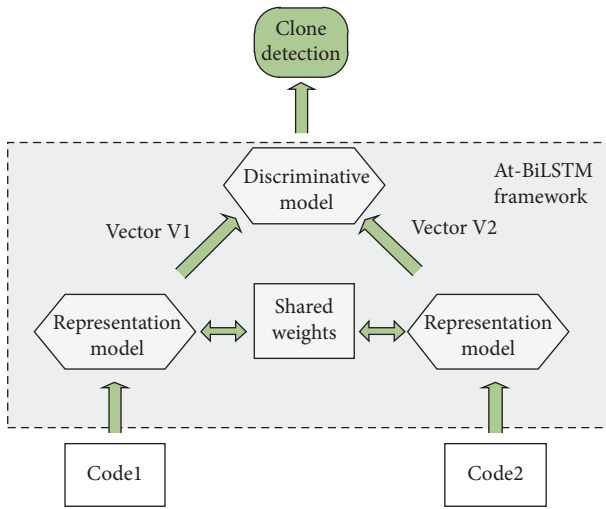


FIGURE 3: The architecture of At-BiLSTM.

Step 1: the source code fragments to be detected are transformed into abstract syntax trees by existing AST tools. For each AST, we decompose it by the granularity of language statement and extract the sequence of statement trees with a preorder traversal. The root of each statement tree is a statement in the corresponding code fragment.

Step 2: each statement tree is transformed into a statement vector via a LSTM encoder with a deep-first traversal algorithm. At that moment, the whole AST representing the source code fragment is transformed into an ordered set of statement vectors.

Step 3: the model generates the final representative vector from a sequence of statement vectors via a specific bidirectional LSTM encoder with the self-attention layer.

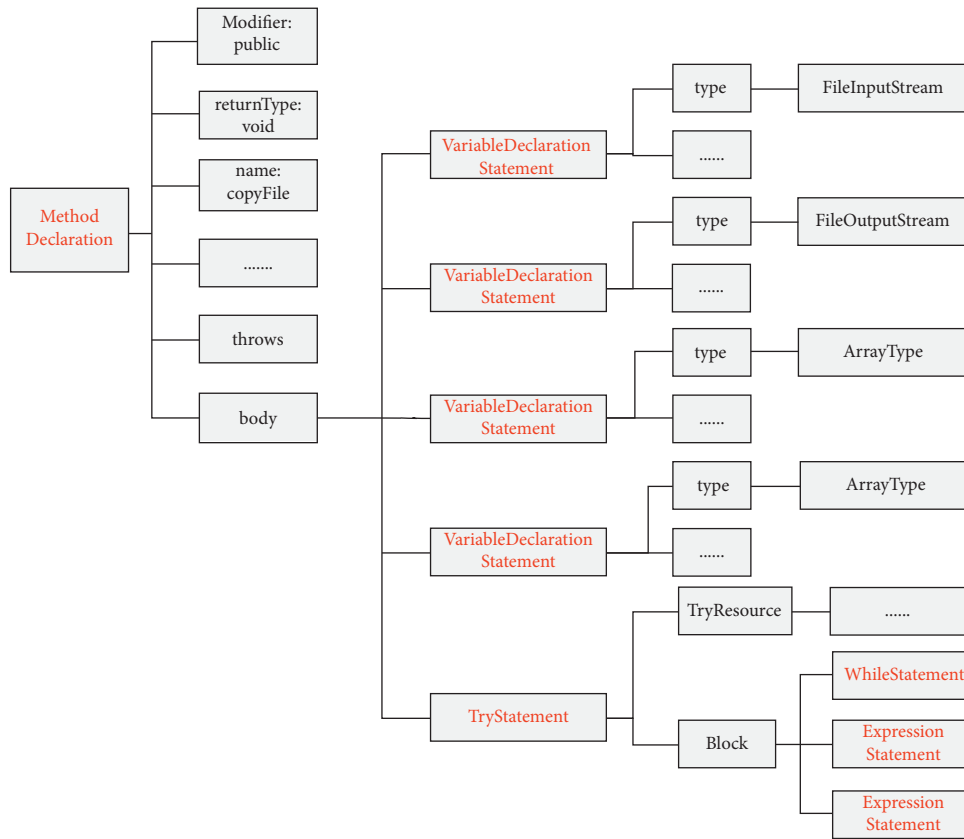
The pair of code vectors generated from the representation model is then loaded into the discriminative model. The discriminator predicts the probability of being clone for the code pair. We can also generate hard labels for the code pair by setting a probability threshold. The detailed design of our model is described in the following sections.

4.2. *Splitting the AST.* Firstly, the source code fragment is transformed to an AST with common parsing tools supplied by language developers, e.g., pycparser [29] and javalang [30]. However, the ASTs g

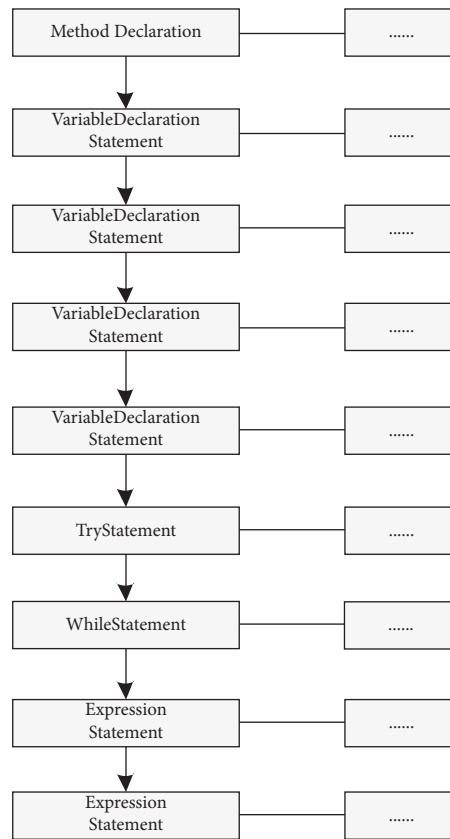
enerated by tools might be too deep to traverse due to the problem of long-term dependency, which reduces the generalization ability of the presentation model. Hence, the strategy to split the whole tree into a sequence of subtrees is adopted in our model to reduce the computational complexity.

There are many methods to decompose the AST into small subtrees without overlapping [31]. In our paper, inspired by the splitting algorithm in [19], we decompose it by the granularity of natural statements. The process of splitting is straightforward: the model firstly scans the AST with a preorder traversal. When it encounters a tree node representing a programming statement, the constructor generates a statement tree taking the node as its root; meanwhile, the new statement tree is added to a FIFO subtree queue. The splitting algorithm based on the preorder traversal ensures that the subtrees generated are arranged in the same way as the statements in the source code. Figure 4 depicts a simplified AST which is generated from a common Java code fragment. Due to the complexity of the AST expression, Figure 4(a) only shows the top 5 layers of the actual AST, and Figure 4(b) represents the execution order of the generated statement trees.

Formally, given an AST  $T$  and a set of statement nodes  $S$ , each statement node  $s \in S$  in  $T$  corresponds with a statement of source code, marked in red in Figure 4. The Method Declaration is treated as a special statement node; thus,  $S = S \cup \{\text{Method Declaration}\}$ . The compound statements with blocks are split into small statement nodes, each of which generates a new subtree. We define the set of separate nodes  $P = \{\text{block, body}\}$ , where block is for splitting the header and body of compound statements such as While and Try Statements, and body for the method declaration. We define  $D(s)$  as the descendants of the statement node  $s \in S$ , and define  $\text{Sub } S(s)$  as the set of substatement nodes of the statement node  $s$ . For any  $d \in D(s)$ , if there exists a path from  $s$  to  $d$  through a node  $p \in P$ , it means the node  $d$  is included by one statement in the body of statement  $s$ . We say that  $d \in \text{Sub } S(s)$  is a substatement node of  $s$ .



(a)



(b)

FIGURE 4: The decomposition of an AST. (a) AST and statement trees. (b) Statement naturalness.

Here, we give the formal definition of a statement tree. The statement tree rooted by the statement node  $s \in S$  in the tree is a syntactic tree consisting of node  $s$  and all of its descendants  $D(s)$  excluding its substatement nodes  $\text{Sub } S(s)$  in  $T$ . For instance, the first statement tree in Figure 4(a) is a subtree rooted by `MethodDeclaration`, including descendant nodes' `Modifier`, `returnType`, `throws`, and `body`. It excludes the nodes `variableDeclarationStatement`, `TryStatement`, and their descendants. According to the splitting algorithm the sequence of statement trees in Figure 4(b) is generated from the whole AST in Figure 4(a).

**4.3. Encoding the Statement Trees.** Recently, people tend to encode the AST with various recursive neural networks [12, 13]. The recursive encoder generates representative vectors when it traverses the syntactic tree from leaf nodes to the root node layer by layer in a bottom-up way. The final vector generated from the root is considered to be the representation of the whole AST, containing the syntactic information of the source code. In order to realize the recursive encoder, people have to transform the original AST into a complete binary tree, which will slightly change the structure. What is worse, the transformation further enlarges the size of AST, which may cause the model to collapse during training. Even if the loss function eventually converges during training, the classification performance of the model is probably affected by the long-term dependence. Inspired by the algorithms used in the field of code comment generation, we encode the statement tree with a simple LSTM network in our approach.

We use a simple encoding method based on a LSTM network in our approach. Each statement tree is encoded with a depth-first traversal algorithm, which is described in the following:

Step 1 (generating the token dictionary and node encoder): we firstly transform all the code fragments in the training set into ASTs and then generate a token dictionary by extracting all the nodes from the ASTs. Finally, we use the word2vec to train a node encoder  $E$  based on the token dictionary with unsupervised learning.

Step 2 (encoding the statement tree nodes): for the given statement tree ST, we encode all the internal nodes in ST with the encoder  $E$ .

Step 3 (generating an ordered set of node vectors): we traverse ST with a deep-first algorithm and then place all the node vectors into an ordered set  $V$  sequentially.

Step 4 (encoding the statement tree): all the vectors in  $V$  are input into a LSTM encoder sequentially, and the output of the LSTM is the vector representation of ST. The parameters in the LSTM encoder, together with other parameters in At-biLSTM, are trained by supervised learning on the training set.

**4.4. Representing the Sequence of Statement Vectors.** Based on the algorithms of Sections 4.2 and 4.3, the AST representing the given code fragment is now transformed to

an ordered vector set  $V_t = \{v_1, v_2, v_3, \dots, v_t\}$  with a weights shared encoder, where  $t$  denotes the number of statement trees and  $v_t$  denotes a vector encoded from the given statement tree.  $v_1, v_2, v_3, \dots, v_t$  are aligned according to the tree traversal order in Section 4.2.

In this paper, we exploit the LSTM network to track the naturalness of statements due to the sequence of statement tree vectors. Figure 5 shows the structure of representation model. The complicated representation model can be divided into 3 layers. The first layer in the bottom is the statement tree encoder, which is composed of a LSTM network with shared weights. Each substatement tree is transformed into a vector in this encoder sequentially. The second layer is a mixed layer which is composed of a bi-directional LSTM network. All the subtree vectors are placed into the mixed layer and transformed into hidden vectors  $h$ . The third layer is an attention layer, which summarizes all the hidden vectors with corresponding attention scores. The encoding process is detailed in the following paragraphs.

The elements  $v_1, v_2, v_3, \dots, v_t$  in  $V_t$  are sequentially placed into a LSTM network. In order to further strengthen the model's ability to capture sequential information, we employ a bidirectional LSTM, where the hidden states of both directions are concatenated to form the new states as follows:

$$\begin{aligned} \vec{h}_t &= \overrightarrow{\text{LSTM}}(v_t), \quad t \in [1, N], \\ \overleftarrow{h}_t &= \overleftarrow{\text{LSTM}}(v_t), \quad t \in [N, 1], \\ h_t &= \left[ \vec{h}_t, \overleftarrow{h}_t \right], \quad t \in [N, 1]. \end{aligned} \quad (5)$$

However, the effect of each statement is different in high-level programming languages, e.g., the statements dealing with business functionality are much more important than common declaration statements. In this paper, we attempt to introduce the self-attention algorithm in light of variant contribution of each statement to the system functionality. To the best of our knowledge, it is the first time to use attention mechanism in the field of code clone. Since our model only includes the encoder, we decide to use the self-attention algorithm. The attention scores can be calculated in many ways as there are various definitions of similarity function. In our paper, equations (6) and (7) are used to calculate the attention scores, which were defined in [29]:

$$\alpha_i = \frac{\exp(h_i^T K)}{\sum_i \exp(h_i^T K)}, \quad (6)$$

$$C = \sum_i \alpha_i h_i, \quad (7)$$

where  $h_i$  denotes the hidden states of the bidirectional LSTM and  $K$  denotes a context vector which is initialized randomly. The author [29] indicates that the inner product of  $h_i$  and  $K$  denotes the contribution of  $h_i$  to the source code vector. The value of  $K$  is updated continuously with other parameters via supervised learning.  $\alpha_i$  denotes the  $i$ th attention score,



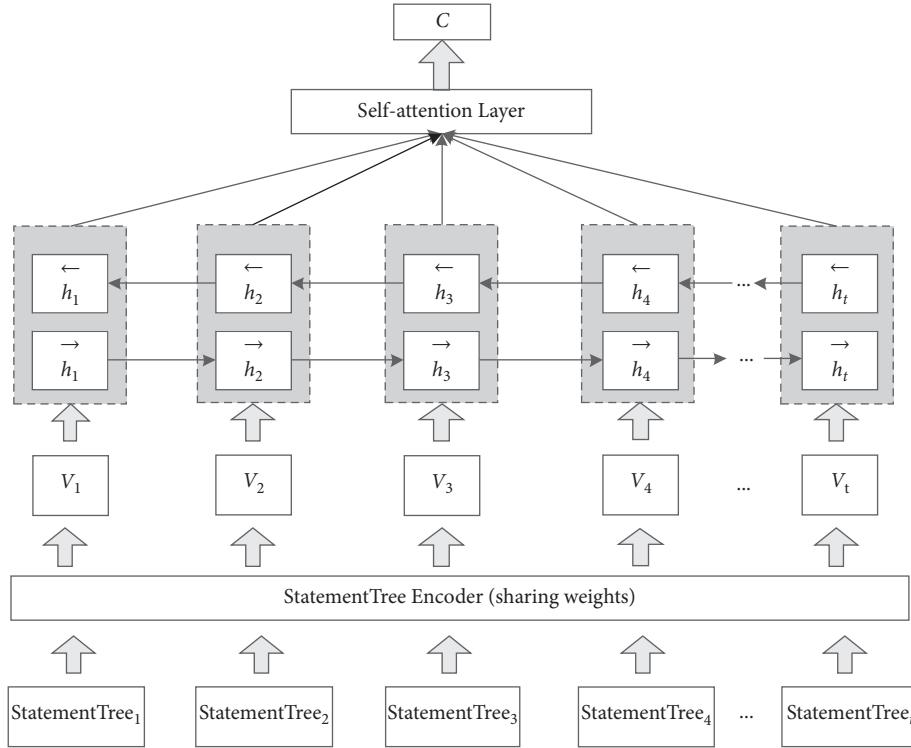


FIGURE 5: Representation model with self-attention.

corresponding with the hidden state  $h_i$ .  $C$  is the representative vector of the given code fragment, which is generated by the weighted average of all the hidden state vectors. Due to the structure of LSTM and self-attention, the vector  $C$  contains the sequential information of all the statements; meanwhile, it intensifies the contribution of key statements. The upper part of Figure 5 shows the encoding process with the self-attention layer.

**4.5. Loss Function and Similarity Computation.** The source code pair is mapped to two corresponding vectors in high-dimensional space via the representation model. We then compute the similarity of these vectors in the discriminator of At-BiLSTM. In this paper, our model is trained with supervised learning as the performance of supervised learning is much better than unsupervised learning. As a typical problem of the binary classification, binary cross-entropy is used as the lost function, which is defined in

$$L = \sum [y \ln \hat{y} + (1 - y) \ln (1 - \hat{y})], \quad (8)$$

where  $y$  is the real probability of code clone and  $\hat{y} \in (0, 1)$  is the prediction probability of our classifier. As the granularity of clone is measured according to the similarity of code pair in the model, we use the method [32] to compute the code pair similarity. Suppose there are two code fragment vectors  $v_1$  and  $v_2$ , which are generated via our representation model, and their distance is defined as  $r = |v_1 - v_2|$  for semantic relatedness. The probability of code pair similarity is computed as follows:

$$\hat{y} = \text{sigmoid}(W_o(r) + b_o), \quad (9)$$

where  $W_o$  is the weight matrix and  $b_o$  is the bias in the discriminative model. In the process of detection, we carefully define a proper threshold  $\tau$ . The detection system identifies the code pair as a clone pair if  $\hat{y} > \tau$ . In this paper, we set  $\tau$  to 0.5 via experiments. The code pairs are identified as clone if the similarity threshold is above 0.5.

**4.6. Evaluation Metrics.** The code clone detection is a typical binary classification problem in machine learning, so we define and calculate the following metrics to evaluate the effectiveness of any clone detection approach. True positive (TP) refers to the number of clone pairs that are identified as clone by the classification system. False positive (FP) refers to the number of nonclone pairs that are identified as clone. True negative (TN) refers to the number of nonclone pairs that are identified as nonclone. False negative (FN) refers to the number of clone pairs that are identified as nonclone. The goal of the clone detection system is to find the proper algorithm with the highest Precision, Recall, and F1 score.

**Precision:** it measures among all of the clone pairs reported by a clone detection approach, how many of them are actually true clone pairs:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (10)$$

**Recall:** it measures among all known true clone pairs, how many of them are detected by a clone detection approach:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (11)$$

F1 score: it combines Precision and Recall to measure the overall accuracy of clone detection. The detection performance is better if the F1 score is higher:

$$F_1 = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (12)$$

## 5. Experiments

*5.1. Dataset Description.* There are very few public datasets for code clone detection as it is very difficult to collect and identify clone pairs. We conduct our experiments on two public datasets BigCloneBench [17] and OJClone [24], which are frequently used to evaluate approaches of detecting code clones. BigCloneBench was released by Svajlenko et al. in 2015. It is the first big-data-curated benchmark of real clones to evaluate the performance of detecting models. According to the authors, the dataset was built by mining clones of frequently implemented functionalities from 25,000 Java projects. Based on the syntactical similarity which was measured as the ratio of lines or tokens that a code fragment shared with another after normalization, the clone pairs in BigCloneBench were divided into 5 groups: type 1, type 2, midtype 3, strong type 3, and type 4. Table 1 summarizes the data distribution in the first version of BigCloneBench in terms of clone types. We conclude from the table that more than 98% clone pairs in BigCloneBench are type 4 clones, which is quite appropriate to evaluate the detection of semantic clones.

Specifically, the clone labels in BigCloneBench were verified by three domain experts, taking 216 hours. The dataset is still been updated, and the current version has about 8 million tagged true clone pairs covering 43 functionalities. However, the authors of recent detection models always conducted experiments on the first version of BigCloneBench, which contains over 6 million true clone pairs and more than 260,000 false clone pairs, covering 10 functionalities. In this paper, we also sample data from the first version of BigCloneBench for comparison.

OJClone is a C language dataset released in 2016 by Mou et al., including 104 programming problems together with various source code solutions which students submit for each problem. There are no labels for code pairs in OJClone as the dataset was designed for code classification in the beginning. However, recent scholars consider any two different source code pairs solving the same programming problem to be T3 or T4 clone, as they realize the same functionality. OJClone is often used to identify T4 clones in recent years.

*5.2. Data Collection.* There are 59,668 code fragments in BigCloneBench, composing more than 6 million true clone pairs and 260,000 nonclone pairs. Due to the huge amount of data in BigCloneBench, we construct the experimental datasets by random sampling. For BigCloneBench, we randomly sample 40,000 pairs of negative samples, 20,000

TABLE 1: Percentage of clone types in BigCloneBench.

| Type 1 | Type 2 | Strong type 3 | Midtype 3 | Type 4 |
|--------|--------|---------------|-----------|--------|
| 0.005  | 0.001  | 0.002         | 0.01      | 0.982  |

pairs of positive samples in midtype 3 and type 4 clones, and 20,000 pairs of positive samples in type 1, type 2, and strong type 3 clones totally.

For OJClone, there are 104 programming problems. Similar to the previous works [12, 19], we select 500 program solutions from each of the first 15 programming problems to obtain 7500 code fragments in OJClone. Table 2 describes the data we collect in the experimental datasets. All the code fragments are parsed into ASTs. The tokens and depths of the ASTs are also shown in Table 2. Lack of domain knowledge, we expect two source code fragments for the same problem belong to a clone pair, and those for various problems are not clone pairs. The 7500 code fragments produce more than 28 million clone pairs, which is extremely time consuming for comparison. Therefore, we randomly select 50,000 samples to construct the experiment dataset for convenience. For each dataset, we randomly divide it into three pieces, of which the proportions are 80%, 10%, and 10% for training, validation, and testing.

*5.3. Experimental Setting.* All the experiments are conducted on a server, having a CPU of 16 cores with GPU acceleration. The model is implemented with a famous deep learning framework, pytorch. In order to generate ASTs from code fragments, we use javalang for java code and pycparser for C code, respectively. For token embedding, we train the token encoder for AST nodes using word2vec with Skip-gram algorithm and set the embedding size to be 128.

The neutral networks for the representation model and discriminative model are all implemented with pytorch. In the encoding of single statement tree, we exploit a common LSTM encoder, where the hidden dimension is set to be 120. In the encoding of statement vector sequence, we exploit a bidirectional LSTM encoder followed by a self-attention layer, where the hidden dimension is set to be 128. We use SGD to train the model and set Batch\_size to 64. We employ the optimizer Adam with a learning rate 0.005 to train our model.

*5.4. Experimental Result.* We compare our model with the existing state-of-the-art neural network approaches by different experiments.

- (i) RAE model [13]: the RAE model transforms the source code pairs into vectors by a recursive encoder, which is trained with unsupervised learning. The model identifies the clone pair by calculating the similarity distance between vectors. With the open-source tool provided by the authors, we use the RAE model to detect clone pairs on our datasets in a supervised learning way.
- (ii) CDLH model [12]: the CDLH model traverses AST with a tree-based LSTM network and then generates

TABLE 2: Statistics of datasets used for clone detection.

| Dataset                       | OJClone | BigCloneBench |
|-------------------------------|---------|---------------|
| Code fragments                | 7500    | 59688         |
| True clone pairs (percentage) | 6.8     | 95.6          |
| Avg tokens                    | 232     | 228           |
| Avg AST depth                 | 12.9    | 9.8           |
| Avg AST nodes                 | 184     | 205           |

the hashed code vectors via a specific hash function for efficiency. It identifies code clone by comparing the hamming distance between hashed vectors. CDLH is not made public by the authors, so we directly cite their results from the paper as their experiments are made on the same datasets.

- (iii) ASTNN model [19]: the ASTNN model is a new code representation tool based on the AST parsing proposed by Zhang et al. in 2019. Their model first decomposes the corresponding ASTs into statement trees and then encodes them, respectively, with a recursive encoder. Finally, through a bidirectional LSTM network with a pooling layer, the statement vectors are transformed into a unified vector representing the given code fragment. The authors indicate that ASTNN is superior to the state-of-the-art models on baseline datasets. Inspired by ASTNN, our model divides the generated AST with the same algorithm. The authors also public their system on GitHub, but we failed to repeat their experiments due to the hardware environment. We directly refer to their experimental result for a contrast in our paper.

Figure 6 shows the comparison of At-biLSTM and three models mentioned above on the OJClone dataset. We see that both the precision and recall rates of RAE and CDLH are relatively poor. It is because that OJClone contains only T3 and T4 clones, and RAE and CDLH are not good at detecting heterogeneous clone pairs. The detection performance of At-BiLSTM and ASTNN are much higher than the former models. At-BiLSTM is better than ASTNN in the aspect of recall rate, while ASTNN is superior in classification precision.

Figure 7 shows the experimental result of At-biLSTM and other models on the BigCloneBench dataset. We see that the classification performance of RAE and CDLH is better on BigCloneBench, as BigCloneBench contains many T1 and T2 clone pairs. Both of RAE and CDLH can detect low-level clone pairs, so their classification precision and recall rates are improved. At-biLSTM and ASTNN once again take the lead in the classification performance. Due to the excellent recall rate, the F1 value of At-BiLSTM is slightly higher than ASTNN, which means At-biLSTM can find more clone pairs in the candidate set.

The big size of generated ASTs, which is shown in Table 2, strengthens the effect of long-term dependency in the process of embedding. The lines of the code (LOC) play an important role in the identification of clone pairs. In BigCloneBench, more than 4.1% code fragments contain more

than 100 statements. We decide to investigate the performance of these classifiers by collecting codes with different lines. Figure 8 shows the classification performance of 4 models according to LOC in BigCloneBench. The X-axis in Figure 8 indicates LOC, which is represented by the lines of the longer code in the given clone pair. The Y-axis indicates the classification performance of the given model. It is clearly evident from Figure 8 that the precision rates of most classifiers reach the peak when the LOC is between 20 and 30. As the LOC increase, the precision rates decrease. In general, the performance of ASTNN and At-BiLSTM are superior to RAE and CDLH. The classifying precision of ASTNN is highest among all the models when LOC is below 50, but it fails to take lead when dealing with complicated codes (LOC > 90). The performance curve of At-BiLSTM is relatively gentle with the increase of LOC. The precision rate of At-BiLSTM overwhelms the other models when LOC is above 70. We guess the attention technique involved in At-BiLSTM has successfully weakened the impact of long-term dependency in the LSTM structure. All the nodes in the long chain are taken into account during embedding with the help of various attention scores.

## 6. Discussion

In this chapter, we take a discussion about the performance of At-BiLSTM and other popular detection models based on the contrast experiments in Section 5. We try to find the key components which affect the classification of At-biLSTM by further experiments.

*6.1. Comparing with Baseline Models.* Compared to other baseline models, we conclude from the experiments that the At-biLSTM model has excellent classification precision and recall rate in code clone detection. The performance of RAE and CDLH is not satisfying since they are both affected by long-term dependence and might lose details in the process of source code representation. ASTNN adopts the strategy of subtree decomposition and recursive encoding, combined with a bidirectional LSTM to leverage the naturalness of statements. The representation model of ASTNN not only reduces the depth of syntax trees but also employs the natural sequence of statements in the code. At-biLSTM decomposes the whole AST into a sequence of statement trees in the same way as ASTNN does. However, we apply a depth-first traversal algorithm in the encoding of statement trees instead of a recursive traversal algorithm which is more complicated in ASTNN. The use of self-attention mechanism in At-biLSTM strengthens the effects of core statements in the clone classification. Experimental results show that At-biLSTM is superior to other baseline models in the detection of type 3 and type 4 clone.

*6.2. The Splitting Granularity of ASTs.* Most of the current approaches [12, 13, 16, 23, 33] based on the AST analysis identify the code clone pairs by traversing the complete AST. In order to evaluate the splitting strategy of At-biLSTM, we conduct contrast experiments on the same dataset. The

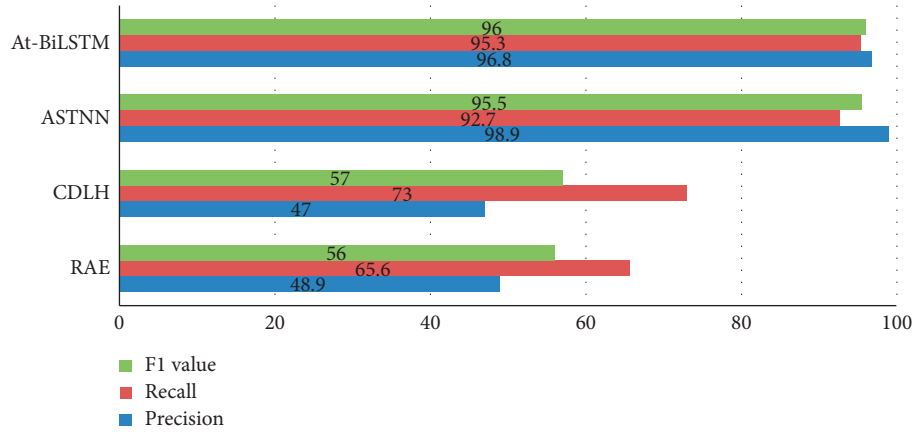


FIGURE 6: Code clone detection models on OJClone.

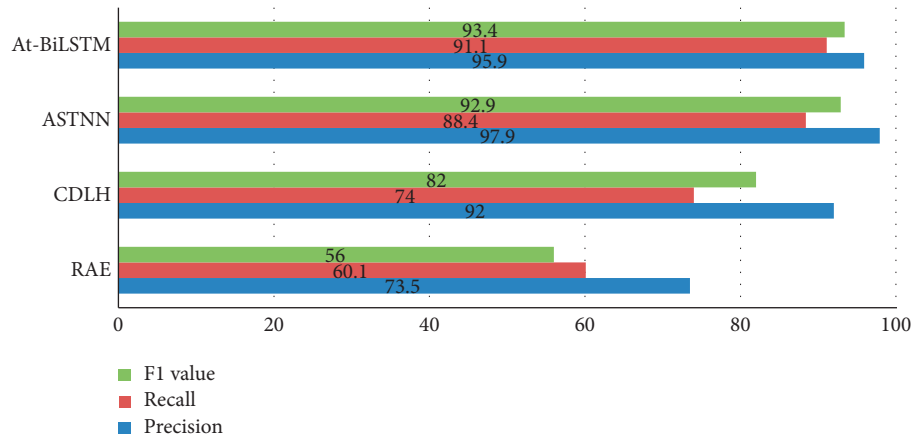


FIGURE 7: Code clone detection models on BigCloneBench.

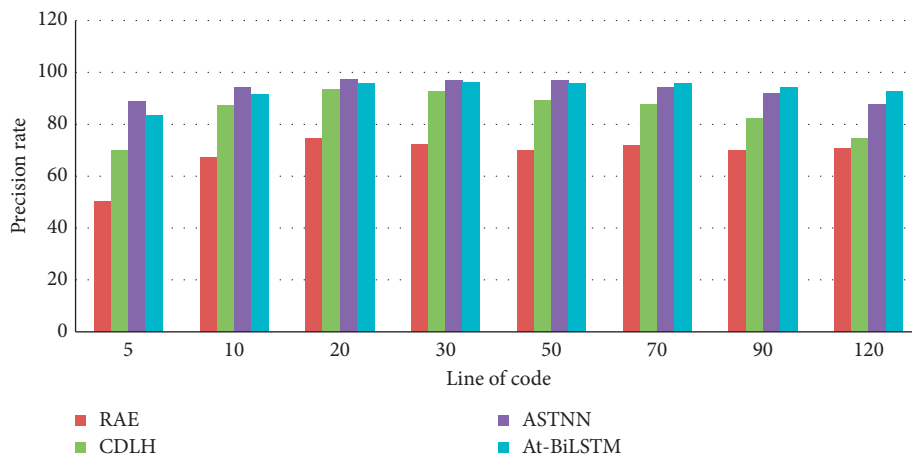


FIGURE 8: Precision of code clone detection models according to LOC.

original AST is decomposed with three different strategies according to the splitting granularity. (1) Model 1: we treat the whole AST as a special statement tree. It is encoded directly by the statement tree encoding algorithm in Section 4.3 without any splitting, and then the classifier predicts the

code clone by comparing the similarity of AST vectors without self-attention. (2) Model 2: we use At-biLSTM to identify code clones. (3) Model 3: we split the AST according to blocks (compound statements including multiple statements within the same brace pairs), e.g., the tree rooted by

TABLE 3: Classification performance depending on AST splitting methods (percentage).

| Description         | F1 score (BigCloneBench) | F1 score (OJClone) |
|---------------------|--------------------------|--------------------|
| AST-full (model 1)  | 75.8                     | 68.9               |
| At-BiLSTM (model 2) | 94.3                     | 95.8               |
| AST-block (model 3) | 93.2                     | 91.5               |

TABLE 4: Clone detection models with various self-attention layers (percentage).

| Description                             | F1 score (BigCloneBench) | F1 score (OJClone) |
|---|--------------------------|--------------------|
| No self-attention                       | 89.6                     | 87.9               |
| Self-attention layer with no parameters | 92.4                     | 91.7               |
| At-BiLSTM                               | 94.3                     | 95.8               |

TryStatement is not divided into subtrees in Model 3. After splitting, the following encoding processes are the same as those in At-biLSTM. Table 3 shows the classification results of these models.

Through the above experimental results, it can be observed that model 1 shows the poorest performance for classification, as it fails to extract the structural information hidden in the AST. The hierarchy of the original AST is too complicated, leading to the long-term dependency problem in the process of encoding. All the statements in the code fragment are evaluated equally by the classifier without the self-attention layer. Model 3 treats the compound statement as an independent statement tree instead of making a further division. The performance of model 3 is close to At-biLSTM. We guess that At-biLSTM achieves the best performance because the splitting of our model better retains the semantic information of the AST.

**6.3. The Use of Self-Attention.** The use of self-attention layer, which stresses the effect of key statements in source clone detection, is a great innovation in our paper. In this section, we investigate the effect of self-attention layers on the classification performance by experiments. We collect data from three independent experiments, as shown in Table 4. The structures of these detection models are the same except the self-attention layers. The first model does not include the self-attention layer, and the vectors representing the code fragments are directly generated with a bidirectional LSTM network. The structure of the second classifier is the same with At-biLSTM, except that it computes the self-attention scores in another way, as shown in equation (13). It is another common calculation for self-attention scores, which is relatively simple without the introduction of other parameters in the model. The attention scores are calculated directly according to equation (13) without the training of parameters  $K$ . The rest part of the second model is the same as At-biLSTM except that all the attention scores of hidden vectors are computed according to equation (13). The third model is At-biLSTM, which trains the self-attention parameters, together with other parameters of the model in a supervised way.

Table 4 shows that the self-attention layer greatly improves the classification performance of the model. The contributions of different statements are various in high-

level programming languages. The core statements implementing the logical functionality are more important than the normal ones. The encoder with self-attention layers can strengthen the effectiveness of core statements in classification and weaken the impact of common statements such as variable declaration. The last two rows of Table 4 are the F1 scores of models implemented with two different self-attention algorithms, from which we find that the classification performances are slightly different. The performance of At-BiLSTM is higher when involving an independent parameter  $K$  in equation (6). We suspect that the classification performance of the second model is slightly affected due to the lack of some adjustments to the input when calculating the attention scores:

$$\alpha_i = \frac{\exp(h_i^T h_i)}{\sum_i \exp(h_i^T h_i)}. \quad (13)$$

**6.4. The GRU.** In the representation model of At-BiLSTM, we use LSTM by default. We have also made experiments by replacing LSTM with Gated Recurrent Units (GRU) in our model. Figure 9 shows the classification performance and training time of our model implemented with LSTM and GRU respectively in the dataset BigCloneBench. We find the GRU model has a slightly poor but comparative performance. However, the training of the GRU model is more efficient than the LSTM model as the structure of GRU is simplified according to LSTM. The researchers have to make a trade-off between the classifier’s precision and efficiency.

The threats to the validity stem from the datasets used in our experiments. The most popular datasets employed in the area of code clone detection are BigCloneBench and OJClone, as the clone data are difficult to accumulate. The data labels in BigCloneBench are considered to be correct, all of which are inspected manually. However, there are only 10 kinds of clone data in BigCloneBench, which are much fewer than those existing in the real world. The data in OJClone was collected for program classification in the beginning. Lack of domain knowledge, we cannot guarantee the programming solutions to the same question are of the same clone class, nor the answers to various questions are of different clone classes.

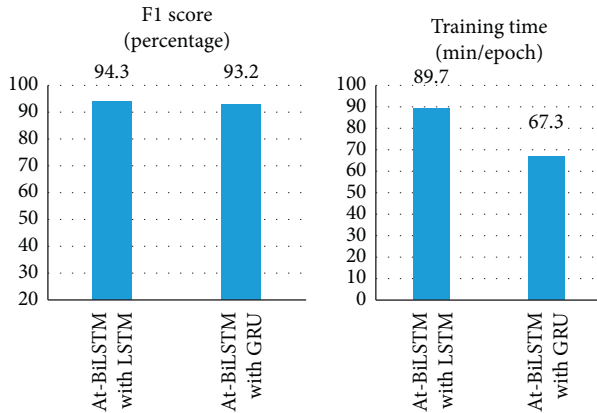


FIGURE 9: Performance of At-BiLSTM with LSTM and GRU on BigCloneBench.

## 7. Conclusion

In this paper, we have presented a novel source code clone detection model At-BiLSTM based on deep learning with self-attention mechanism, which successfully captures both the syntactic and semantic information of the code in the process of encoding. At-BiLSTM contains a representation model and a discriminative model. The contributions of core statements are deeply enhanced with the introduction of self-attention mechanism. Contrast experiments show that our model significantly outperforms most of the existing neural network models, especially in the detection of type 3 and type 4 clones.

We will focus on two aspects in the future. Firstly, the At-BiLSTM should be used to detect large-scale code clone. At present, both the algorithms for clone detection and data collected in the labs cannot be applied in the real world. We will study and improve the performance of At-BiLSTM in industry. Secondly, the source code representation model embedded in At-BiLSTM will be applied in different areas of software engineering, as the representation algorithm is general and cross platform. For future work, we plan to extend our representation algorithm to new applications in software engineering such as source code classification, code completion, code comment generation, and deep code search.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the National Key Research and Development Program of China under Grant no. 2019QY1300.


## References

- [1] G. Antoniol, U. Villano, E. Merlo, and M. Di Penta, "Analyzing cloning evolution in the Linux kernel," *Information and Software Technology*, vol. 44, no. 13, pp. 755–765, 2002.
- [2] T. Kamiya, S. Kusumoto, and K. Inoue, "CCFinder: a multilingual token-based code clone detection system for large scale source code," *IEEE Transactions on Software Engineering*, vol. 28, no. 7, pp. 654–670, 2002.
- [3] K. Hotta, Y. Sano, Y. Higo, and S. Kusumoto, "Is duplicate code more frequently modified than non-duplicate code in software evolution?: an empirical study on open source software," in *Proceedings of the Joint ERCIM Workshop on Software Evolution (EVOL) and International Workshop on Principles of Software Evolution (IWPSE)*, Antwerp, Belgium, September 2010.
- [4] J. Krinke, "A study of consistent and inconsistent changes to code clones," in *Proceedings of the 14th Working Conference on Reverse Engineering (WCRE 2007)*, Vancouver, BC, Canada, October 2007.
- [5] S. Bellon, R. Koschke, G. Antoniol, J. Krinke, and E. Merlo, "Comparison and evaluation of clone detection tools," *IEEE Transactions on Software Engineering*, vol. 33, no. 9, pp. 577–591, 2007.
- [6] H. Sajjani, V. Saini, J. Svajlenko, C. K. Roy, and C. V. Lopes, "SourcererCC," in *Proceedings of the 38th International Conference on Software Engineering-ICSE'16*, Austin, TX, USA, May 2016.
- [7] R. Tairas and J. Gray, "An information retrieval process to aid in the analysis of code clones," *Empirical Software Engineering*, vol. 14, no. 1, pp. 33–56, 2009.
- [8] Y. Liu, D. Poshyvanyk, R. Ferenc, T. Gyimóthy, and N. Chrisochoides, "Modeling class cohesion as mixtures of latent topics," in *Proceedings of the IEEE International Conference on Software Maintenance*, Edmonton, AB, Canada, September 2009.
- [9] D. Scott, S. T. Dumais, and G. W. Furnas, "Indexing by latent semantic analysis," *Journal of the Association for Information Science & Technology*, vol. 1, pp. 2–9, 2010.
- [10] W. Kong and W. J. Li, "Isotropic hashing," in *Proceedings of the International Conference on Neural Information Processing Systems*, Lake Tahoe, NV, USA, December 2012.
- [11] H. Yu, W. Lam, L. Chen, G. Li, T. Xie, and Q. Wang, "Neural detection of semantic code clones via tree-based convolution," in *Proceedings of the 2019 IEEE/ACM 27th International Conference on Program Comprehension (ICPC)*, pp. 70–80, Montreal, QC, Canada, May 2019.
- [12] H. Wei and L. Ming, "Supervised deep features for software functional clone detection by exploiting lexical and syntactical information in source code," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, Melbourne, Australia, August 2017.
- [13] M. White, M. Tufano, C. Vendome, and D. Poshyvanyk, "Deep learning code fragments for code clone detection," in *Proceedings of the 31st IEEE/ACM International Conference on Automated Software Engineering-ASE 2016*, Singapore, September 2016.
- [14] J. Mayrand, C. Leblanc, and E. Merlo, "Experiment on the automatic detection of function clones in a software system using metrics," in *Proceedings of the 1996 International Conference on Software Maintenance (ICSM'96)*, Monterey, CA, USA, November 1996.
- [15] S. Lee, "SDD: high performance code clone detection system for large scale source code," in *Proceedings of the Companion*

- to the ACM Sigplan Conference on Object-Oriented Programming, San Diego, CA, USA, October 2005.
- [16] L. Jiang, "DECKARD: scalable and accurate tree-based detection of code clones," in *Proceedings of the ICSE*, pp. 96–105, Minneapolis, MN, USA, May 2007.
  - [17] J. Svajlenko and C. K. Roy, "Evaluating clone detection tools with BigCloneBench," in *Proceedings of the 2015 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, Bremen, Germany, September 2015.
  - [18] T. Mikolov, G. s. Corrado, K. Chen, and J. Dean, *Efficient Estimation of Word Representations in Vector Space*, 2013, <http://arxiv.org/abs/1301.3781>.
  - [19] J. Zhang, X. Wang, H. Zhang, H. Sun, K. Wang, and X. Liu, "A novel neural source code representation based on abstract syntax tree," in *Proceedings of the 2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE)*, ACM, Montreal, Canada, May 2019.
  - [20] D. Tang, B. Qin, and T. Liu, "Document modeling with gated recurrent neural network for sentiment classification," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, Portugal, September 2015.
  - [21] R. Komondoor and S. Horwitz, "Using slicing to identify duplication in source code," *Static Analysis*, pp. 40–56, 2001.
  - [22] J. Krinke, "Identifying similar code with program dependence graphs," in *Proceedings of the 8th Working Conference on Reverse Engineering*, Stuttgart, Germany, October 2001.
  - [23] L. Li, F. He, W. Zhuang, M. Na, and B. Ryder, "CCLearner: a deep learning-based clone detection approach," in *Proceedings of the 2017 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, Shanghai, China, September 2017.
  - [24] L. Mou, L. Ge, J. Zhi, Z. Lu, and W. Tao, "Convolutional neural network over tree structures for programming language processing," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, Phoenix, AZ, USA, February 2016.
  - [25] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, <http://arxiv.org/abs/1409.0473>.
  - [26] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
  - [27] H. Liang, J. Zou, K. Zuo, and M. J. Khan, "An improved genetic algorithm optimization fuzzy controller applied to the wellhead back pressure control system," *Mechanical Systems and Signal Processing*, vol. 142, Article ID 106708, 2020.
  - [28] Z. Liu, B. Hu, B. Huang, L. Lang, H. Guo, and Y. Zhao, "Decision optimization of low-carbon dual-channel supply chain of auto parts based on smart city architecture," *Complexity*, vol. 2020, no. 5, 14 pages, Article ID 2145951, 2020.
  - [29] "pyparser," <https://pypi.python.org/pypi/pyparser>.
  - [30] "javalang," <https://github.com/c2nes/javalang>.
  - [31] H. Liang, D. Zou, Z. Li, K. M. Junaid, and Y. Lu, "Dynamic evaluation of drilling leakage risk based on fuzzy theory and PSO-SVR algorithm," *Future Generation Computer Systems*, vol. 95, pp. 454–466, 2019.
  - [32] K. S. Tai, R. Socher, and C. D. Manning, "Improved semantic representations from tree-structured long short-term memory networks," *Computer Science*, vol. 5, no. 1, p. 36, 2015.
  - [33] C. Xu, "A novel recommendation method based on social network using matrix factorization technique," *Information Processing & Management*, vol. 54, no. 3, pp. 463–474, 2018.

## Research Article

# Photovoltaic Generation Prediction of CCIPCA Combined with LSTM

E. Zhu <sup>1,2</sup> and D. Pi<sup>2</sup>

<sup>1</sup>College of Internet of Things Engineering, Jiangsu Vocational College of Information Technology, Wuxi 214153, China

<sup>2</sup>College of Computer Science and Technology, Nanjing University of Aeronautics & Astronautics, Nanjing 211106, China

Correspondence should be addressed to E. Zhu; [erxi666@163.com](mailto:erxi666@163.com)

Received 4 May 2020; Revised 25 August 2020; Accepted 4 September 2020; Published 15 September 2020

Academic Editor: Zhihan Lv

Copyright © 2020 E. Zhu and D. Pi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to remedy problems encompassing large-scale data being collected by photovoltaic (PV) stations, multiple dimensions of power prediction mode input, noise, slow model convergence speed, and poor precision, a power prediction model that combines the Candid Covariance-free Incremental Principal Component Analysis (CCIPCA) with Long Short-Term Memory (LSTM) network was proposed in this study. The corresponding model uses factor correlation coefficient to evaluate the factors that affect PV generation and obtains the most critical factor of PV generation. Then, it uses CCIPCA to reduce the dimension of PV super large-scale data to the factor dimension, avoiding the complex calculation of covariance matrix of algorithms such as Principal Component Analysis (PCA) and to some extent eliminating the influence of noise made by PV generation data acquisition equipment and transmission equipment such as sensors. The training speed and convergence speed of LSTM are improved by the dimension-reduced data. The PV generation data of a certain power station over a period is collected from SolarGIS as sample data. The model is compared with Markov chain power generation prediction model and GA-BP power generation prediction model. The experimental results indicate that the generation prediction error of the model is less than 3%.

## 1. Introduction

With the gradual implementation of the “Internet + energy” policy, the PV generation industry is rapidly developing. The proposed incorporation of artificial intelligence has instilled new incentives for the PV generation industry. PV generation is susceptible to influence from extreme weather, distorting the predicted results and adding trouble in dispatching when the power system is connected to the grid [1], which may even lead to power resource wastage. Numerous methods in artificial intelligence exist for photovoltaic generation prediction such as support vector machines [2], Markov chains [3], and neural networks. Neural networks are widely used in PV generation prediction, and a number of power generation prediction models have surfaced to this effect, including neural networks based on time series [4], GA-BP neural network [5], deep belief network [6], and fuzzy neural

network [7]. Such models find it difficult to overcome issues in convergence and local optimization, and they suffer from large prediction errors, mainly because many network input compositions are present, along with noises in the sample data. Various scholars have used principal component analysis (PCA) [8] and other algorithms to preprocess the collected data; however, PCA and other algorithms find it difficult to process large amounts of data collected in real time due to their own defects. In view of the above problem, this paper adopted CCIPCA, which is suitable in analysing large data sets to extract key data affecting power generation prediction, to reduce dimensions in data and eliminate noise. Simultaneously, combined with LSTM [9], a power generation prediction model was established to overcome problems such as local convergence and slow training speed, controlling the prediction error of PV generation within 3%.



## 2. Relevant Works

**2.1. Analysis of Correlation Factors in Power Generation Prediction.** The sample data adopted the data of PV station no. 11282, Zhangdian District, Zibo City, Shandong Province, China, which was provided by SolarGIS. Its geographical location is 118 degrees east longitude and 32 degrees north latitude. The data includes total solar radiation data such as horizontal radiation GHI and normal direct radiation DNI and meteorological parameters like temperature, humidity, and pressure, as well as environmental parameters such as elevation, surface inclination angle, and surface azimuth angle. There were a total of 37 dimensions and 30 minutes of resolution.

Considering that parameters like geographical location are constant and have little influence on power generation prediction, the prediction model will not be considered. Meteorological and historical power generation data were taken as the main influencing factors of the model, including power generation every 30 minutes, power generation at the moment in history, environmental temperature, environmental humidity, wind speed, wind direction, radiation amount, and other indicators. According to formula (1), the correlation analysis of the evaluation factors was carried out:

$$\zeta = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}} \quad (1)$$

In Figure 1, the linear features denoted by (a) and (b) are consistent, which are either a line segment or a part of an object without any difference. However, given the overall information, the information represented by (c) is completely inconsistent. Thus, the question of how to extract this nonlinear feature effectively is raised. This question is determined by the micro network structure embedded in NIN network, that is, a full connection layer consisting of two layers of convolution. In the neural network, two-layer fully connected hidden neurons are capable of approximating arbitrary curves, where  $\zeta$  is the correlation coefficient,  $x_i$  is the  $i$ -th value of factor  $x$ ,  $\bar{x}$  is the mean of factor  $x$ ,  $y_i$  is the  $i$ -th value of factor  $y$ , and  $\bar{y}$  is the mean of factor  $y$ . Table 1 shows the correlation coefficient analysis of PV generation with meteorological factors and historical power generation.

Table 1 demonstrates that electricity generation is highly correlated with solar radiation and historical electricity generation, valued at 0.92 and 0.86, respectively. Moreover, it had a moderate correlation with environmental temperature of 0.56 and had a weak correlation with wind direction and wind speed, though it was found to be negatively correlated with environmental humidity. Therefore, the prediction model adopted environmental temperature, environmental humidity, solar radiation, and historical power generation as its input components, and the correlation between various input components was observed to be strong; hence, carrying out a principal component analysis was required.

**2.2. CCIPCA.** Due to the advent of big data in PV generation, data preprocessing is particularly important. The PV generation system possesses a simple structure; however, it contains a large amount of equipment and has many failure-prone points. Additionally, the collected data has noise, which introduces obstacles in processing big data. The traditional principal component analysis (PCA) reduces dimensions by eliminating data in the dimension having small variance, which maximizes the information of the original data and removes noise to a certain extent. However, PCA must input all sample data before starting the analysis, which does not align with the objectives of big data; hence, the Candid Covariance-free Incremental Principal Component Analysis (CCIPCA) method was proposed. Unlike the batch method, which uses the covariance matrix to calculate eigenvalues and eigenvectors, CCIPCA initially eliminates the calculation of the covariance matrix using the asymptotic method to estimate the principal component values obtained by approximating the batch method. This data processing technique satisfies the requirements of photovoltaic data processing.

Recently, research related to incremental principal component analysis has been ongoing. Oja and Karhunen et al. proposed the SGA algorithm [10], where, after receiving new sample data, the algorithm continuously refreshes the estimated value of feature vectors and normalizes the new estimated value with all of the previous feature vectors. Sanger put forward the GHA algorithm, which uses a single-layer linear neural network to extract principal elements [11]. Furthermore, Weng et al. described the CCIPCA algorithm [12], which is comprised of the following framework: gradually obtain data  $x_1, x_2, \dots, x_n, \dots$  and calculate the previous  $k$  principal elements  $v_{1n}, v_{2n}, \dots, v_{kn}$ . The calculation process of CCIPCA is given below.

According to derivation of PCA by the maximum variance theory [13], the total variance  $A$  of the data set  $X$  projected along a unit vector  $w$  is obtained as follows:

$$A = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T = \frac{1}{n} \sum_{i=1}^n u_i u_i^T, \quad (2)$$

where  $\bar{x}$  denotes the mean of data set  $X$ ; that is  $\bar{x} = ((\sum_{i=1}^n x_i)/n)$ ,  $u_i = x_i - \bar{x}$ , and  $A$  denotes the covariance matrix of data set  $X$  (the general covariance matrix is divided by  $n - 1$ , where  $n$  is used).

The calculation formula of the  $i$ -th eigenvalue and the  $i$ -th eigenvector is as follows:

$$\lambda_{in} w_{in} = A w_{in}, \quad (3)$$

where  $w_{in}$  denotes the  $n$ -th eigenvector to be calculated at the  $i$ -th input and  $\lambda_{in}$  denotes the corresponding eigenvalue. In order to speed up the iteration of CCIPCA algorithm, the whole iteration is based on the product  $\lambda_{in} w_{in}$  of eigenvalues and eigenvectors. When the  $i$ -th data is input, the following formula is obtained:

$$v_{in} = \lambda_{in} w_{in} = A w_{in}. \quad (4)$$

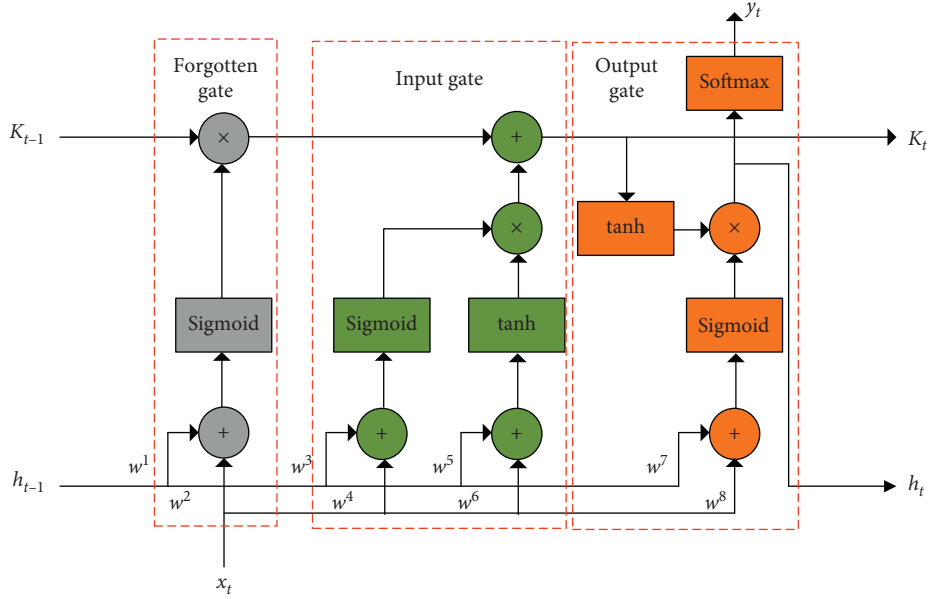


FIGURE 1: Cell structure of LSTM.

TABLE 1: Correlation coefficient between PV generation and other factors.

| Factors | ET    | EH    | WS    | WD    | RQ    | HQ    | PG    |
|---------|-------|-------|-------|-------|-------|-------|-------|
| ET      | 1.00  | -0.42 | 0.12  | 0.02  | 0.48  | 0.55  | 0.56  |
| EH      | -0.42 | 1.00  | -0.17 | -0.33 | -0.36 | -0.24 | -0.38 |
| WS      | 0.12  | -0.17 | 1.00  | -0.21 | 0.22  | 0.12  | 0.22  |
| WD      | 0.02  | -0.33 | -0.21 | 1.00  | 0.06  | 0.20  | 0.07  |
| RQ      | 0.48  | -0.36 | 0.22  | 0.06  | 1.00  | 0.96  | 0.92  |
| HQ      | 0.55  | -0.24 | 0.12  | 0.20  | 0.96  | 1.00  | 0.86  |
| PG      | 0.56  | -0.38 | 0.22  | 0.07  | 0.92  | 0.86  | 1.00  |

ET: environment temperature; EH: environment humidity; WS: wind speed; WD: wind direction; RQ: radiation quantity; HQ: historical quantity; PG: power generation.

Based on formulae (2) and (4),

$$v_{in} = \frac{1}{n} \sum_{i=1}^n u_i u_i^T w_{in}. \quad (5)$$

The product  $v_{in}$  of eigenvalue and eigenvector is obtained by iteration. As the eigenvector is normalized,  $\lambda_{in}$  and  $w_{in}$  can be separately obtained with  $\|v_{in}\|$  and  $(v_{in}/\|v_{in}\|)$ .

Formula (5) uses  $(v_{i,n-1}/\|v_{i,n-1}\|)$  to approximate  $w_{in}$  and replace it. The basic iterative formula of CCIPCA can be obtained by transformation.

$$v_{in} = \frac{n-1}{n} v_{i,n-1} + \frac{1}{n} u_n u_n^T \frac{v_{i,n-1}}{\|v_{i,n-1}\|}, \quad (6)$$

where  $(n-1/n)$  denotes the weight of the previous iteration value  $v_{i,n-1}$  and  $(1/n)$  denotes the adjustment step size of the iteration.

With the adjustment of the iteration vector  $v_{in}$  by the new input data  $u_n$ , that is,  $u_n = x_n - \bar{x}$ ,  $v_{in}$  gradually converges to the  $i$ -th eigenvector. For different  $v_i$ , formula (6) can be used for iteration with the different input vectors  $u_n$ .

The first eigenvector has been obtained by iteration. Firstly,  $u_{1n} = u_n$  is projected onto the previous eigenvector (now the first eigenvector), and the residual data is obtained as follows:

$$u_{2n} = u_{1n} - u_{1n}^T \frac{v_{1n}}{\|v_{1n}\|} \frac{v_{1n}}{\|v_{1n}\|}. \quad (7)$$

$u_{2n}$  denotes the output of the second eigenvector. The third and fourth eigenvector can be similarly obtained. Since the residual data is orthogonal to the data recovered from the previous eigenvector, all orthogonal eigenvectors can be obtained. In addition, the mean should be updated every time when a new data is input. The mean of the  $n$ -th data input adopts the following iterative formula:

$$\bar{x}_n = \frac{n-1}{n} \bar{x}_{n-1} + \frac{1}{n} u_n^T. \quad (8)$$

The CCIPCA solution process is summarized in Algorithm 1.

The CCIPCA algorithm has been widely used in the field of big data processing as well as the decomposition of large matrices due to its good convergence, which has achieved beneficial results. Meanwhile, as a benchmark algorithm, CCIPCA has been cited in various incremental algorithms. In this paper, it was applied in big data preprocessing of the PV station with a time dimension to make up for PCA preprocessing defects.

**2.3. LSTM.** In recent years, the improvement of LSTM has been continuously carried out. Yao et al. [14] proposed an improved NLP method based on Long Short-Term Memory (LSTM) structure, whose parameters are randomly discarded when they are passed backwards in the recursive projection layer, and it is used to overcome practical problems, such as high training complexity, computational

```

Input data:
  original data sequence  $x_1, x_2, \dots, x_n, \dots$ ,
  % the algorithm can pause the output of projection matrix at any time.
  dimension of low-dimensional space  $k$ .
Initialization:  $\bar{x}_0 = 0$ 
Output: projection matrix  $W^* = \{w_1 = (v_1/\|v_1\|), w_2 = (v_2/\|v_2\|), \dots, w_k = (v_k/\|v_k\|)\}$ 
Iteration steps:
  For  $n = 1, 2, \dots$ , % Run the following steps to update  $k$  feature vectors
     $u_{1n} = u_n = x_n - \bar{x}_{n-1}$ ;
    If  $n == 1$ :  $\bar{x}_n = \bar{x}_{n-1} + u_n^T$ 
    Else:  $\bar{x}_n = (n-1/n)\bar{x}_{n-1} + (1/n)u_n^T$ ; % Update mean vector
    For  $i = 1, 2, \dots, \min\{k, n\}$  do:
      (a) If  $i = n$ :  $v_{in} = u_{in}$ ; % Initialize the  $i$ -th eigenvector
      (b) Otherwise:
         $v_{in} = (n-1/n)v_{i,n-1} + (1/n)u_{in}u_{in}^T(v_{i,n-1}/\|v_{i,n-1}\|)$ 
         $u_{i+1,n} = u_{in} - u_{in}^T(v_{in}/\|v_{in}\|)(v_{in}/\|v_{in}\|)$ 

```

ALGORITHM 1: CCIPCA solution process.

difficulties in large-scale content scenarios, high retrieval complexity, and lack of probabilistic significance. Tian et al. [15] proposed a hybrid prediction modelling strategy by combining the autocorrelation local characteristic-scale decomposition and the improved LSTM neural network. Zhang et al. [16] proposed a LSTM approach for bearing performance degradation evaluation. Numerical results show that the proposed LSTM method can effectively predict the remaining service life of the bearing. Li et al. [17] proposed a LSTM method for fault diagnosis and isolation of wind turbine, where stochastic forest algorithm is applied to make decision. Liu et al. [18] combined probabilistic decision-making methods and proposed a Bayesian LSTM algorithm for intelligent fault early-warning of nuclear power machinery. Mirza et al. [19] introduce efficient online learning algorithms based on the Long Short-Term Memory (LSTM) networks that employ the covariance information. They reduce the number of system parameters through the weight matrix factorization, where they convert the LSTM weight matrices into two smaller matrices in order to achieve high learning performance with low computational complexity.

LSTM is an evolutionary version of RNN, which effectively addresses the issue of long-term dependence of effective information in time series and has been broadly applied for different fields. In comparison to other models, the LSTM method is more sensitive to the trivial features in the historical data, easier to capture the details, suitable for big data processing, and more accurate in time series prediction. LSTM replaces the hidden layer of neurons in RNN with a memory unit to record the dependency relationship between time series data, which then rids itself of problems like gradient disappearance and gradient explosion occurring in RNN. LSTM utilizes a “gate” to control information selection. Compared with a dropout operation, this process is not random, though it implements Boolean selection based on sigmoid operation results with “0” signifying the

forgetting of information and “1” meaning the remembering of information. The LSTM structure includes three gate structures that adjust the information flow, namely, the forgotten gate, input gate, and output gate, as shown in Figure 1.

The forgotten gate determines the degree to which the unit state  $K_{t-1}$  at the last time is retained to the current state  $K_t$ , which is calculated as

$$f_t = \text{Sigmoid}(W_f[h_{t-1}, x_t] + c_f), \quad (9)$$

where  $f_t$  is the output value and the bias of the forgotten gate.

The input gate controls the extent to which network input  $x_t$  is saved to the cell state  $K_t$  at the previous time, which is calculated as

$$\begin{aligned} K_t &= f_t c_t + i_t \tilde{K}_t, \\ i_t &= \text{Sigmoid}(W_i[h_{t-1}, x_t] + c_i), \\ \tilde{K}_t &= \tanh(W_k[h_{t-1}, x_t] + c_k). \end{aligned} \quad (10)$$

The output gate controls the current output value  $h_t$  of unit state  $K_t$  outputting to LSTM, which is calculated as

$$\begin{aligned} o_t &= \text{Sigmoid}(W_o[h_{t-1}, x_t] + c_o), \\ h_t &= o_t \tanh(K_t). \end{aligned} \quad (11)$$

If the output value has reached the threshold value required by the memory unit, the product of the output value with the calculated value of the current layer is taken as the output, and the calculation is carried out in the next layer. If the threshold is not reached, the memory unit will forget it.

Different from the traditional RNN, LSTM can solve the issues of gradient disappearance and gradient explosion in the training process, yielding more accurate prediction of a long-term time series. LSTM is developed to address data uncertainty while considering complicated situation of the operation.

### 3. PV Generation Prediction

**3.1. Process of PV Generation Prediction.** Power generation prediction conducted by CCIPCA combined with LSTM conforms to current trends in photovoltaic big data. The prediction process initially processes the CCIPCA algorithm for the collected big data samples, which then establishes the LSTM networks and implements the training of the sample data on the LSTM network. After completion, the sample data is input in order to acquire the prediction results output, as shown in Figure 2. When training big data samples, it is necessary to continuously tune the discarded neuron probability, attenuation rate, and learning rate.

**3.2. Data Source and Preprocessing.** SolarGIS is a solar resource assessment tool developed by SolarGIS S.R.O. in Europe, which uses satellite remote sensing data, GIS technology, and advanced scientific algorithms to obtain a high-resolution database of solar resources and climatic factors. In this paper, data from January 1, 2014, to December 31, 2019, were collected from the SolarGIS database as sample data, of which its volume reached 39.4 PB. Using formula (1), the data within a certain period were used for factor correlation evaluation, and factors with an absolute value coefficient above 0.2 were selected. In addition to selecting highly correlated factors, this paper also required two historical PV power generation horizontal and vertical mean data as the model's input sequences to control the disturbance brought about by extreme weather, as extreme weather is a gradual process that takes place in a short amount of time. The formulae are as follows:

$$\begin{aligned}\bar{Q}_k &= \frac{(Q_{k-1} + \dots + Q_{k-m})}{m}, \\ \bar{Q}'_k &= \frac{(Q'_{k-1} + \dots + Q'_{k-n})}{n},\end{aligned}\quad (12)$$

where  $Q_{k-i}$ ,  $i = 1, 2, 3, \dots, m$ , represents the generating capacity of the previous continuous time units and  $\bar{Q}_k$  represents the average of horizontal generating capacity.  $\bar{Q}'_k$  is the average vertical generating capacity of the same time unit before the previous day and  $Q'_{k-i}$ ,  $i = 1, 2, 3, \dots, n$ , is the electrical generation of the previous  $n$  days.

Actual power generation always fluctuates around the mean curve of the horizontal and vertical power generation, which can be used as a stable factor in the prediction model to eliminate the interference of extreme weather in power generation prediction. Although weather conditions are very sporadic, the seasonal law always changes with Earth's revolution and rotation.

Here, the horizontal mean of the first 10 time units and the vertical mean of the same time unit of the first 5 days were selected as the input. This was done because if the number choices were more, the average value would change significantly, which was not conducive to the stability of the prediction model. If there were fewer number choices, the average value would not change significantly, which was not conducive to measuring the impact of extreme weather

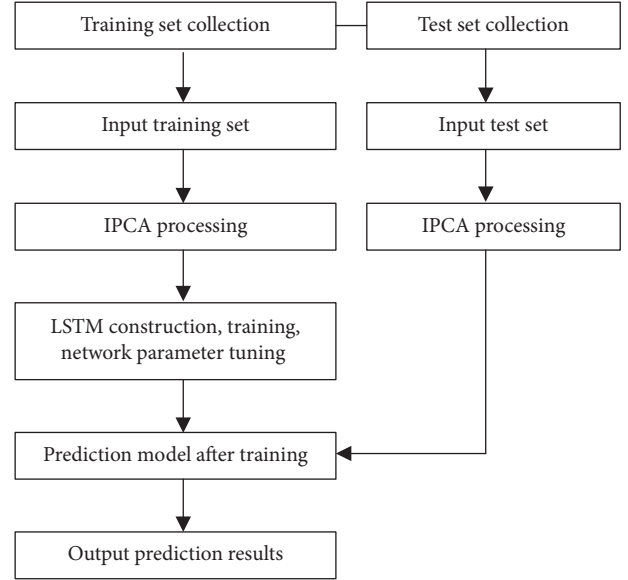


FIGURE 2: Flow chart of power generation prediction by IPCA combined with LSTM.

changes on the prediction model, resulting in inaccuracies of prediction.

After correlation coefficient analysis, the dimension of sample data was reduced to 18. Some sample data are listed in Table 2. After the sample data were sorted, the data was preprocessed with Z-Score standardization to eliminate influences from different feature dimensions and improve the convergence speed of the model. Following completion of the prediction, a reverse operation would be carried out to achieve the real predicted value recovery. Subsequently, CCIPCA was preprocessed to map the sample data to dimensions set by the investigators, such as solar radiation amount and historical power generation. Finally, orthogonalization removed influences in the relationship between noise and dimensions.

Solar radiation adopts two aspects: global horizontal radiation (GHI) and direct normal radiation (DNI). Meteorological parameters are air temperature (TEMP) at 2 m, RH, average WS and WD at 10 m, precipitation RH, and atmospheric pressure.

CCIPCA preprocessing was implemented. 48 pieces of data on June 6, 2019, were used as data series input one by one. The low-dimensional spatial dimension was set to 18. The eigenvector and eigenvalue were calculated, and principal component analysis was conducted to remove the influence of noise and the relationship between dimensions. The results of principal component analysis are shown in Table 3.

It can be determined from Table 3 that the cumulative variance contribution rate of the first six eigenvalues is more than 80%. Therefore, six principal components were selected, and the corresponding eigenvectors of the six principal components were selected to construct the transformation matrix. The original data and the transformation matrix are calculated, and the input variables obtained were input into LSTM for prediction.

TABLE 2: Part of sample data after sorting.

| Date                  | DNI (W. H/m <sup>2</sup> ) | GHI (W. H/m <sup>2</sup> ) | TEMP (°C) | RH (%) | WS (m/s) | WD (°) | AP (Pa) | $\bar{Q}_{10}$ (MWp) | $\bar{Q}_5^d$ (MWp) |
|-----------------------|----------------------------|----------------------------|-----------|--------|----------|--------|---------|----------------------|---------------------|
| 2020-06-06 05: 30: 00 | 11                         | 11                         | 21.4      | 73.2   | 0.3      | 294    | 1000.2  | 0                    | 0.1372              |
| 2020-06-06 06: 30: 00 | 78                         | 80                         | 23.9      | 61     | 0.7      | 349    | 1000.2  | 0.0033               | 0.8562              |
| 2020-06-06 07: 30: 00 | 197                        | 202                        | 26.1      | 53.6   | 0.7      | 266    | 1000.5  | 0.0538               | 2.1322              |
| 2020-06-06 08: 30: 00 | 322                        | 325                        | 28.4      | 47.5   | 1.7      | 232    | 1000.9  | 0.1994               | 3.2794              |
| 2020-06-06 09: 30: 00 | 484                        | 481                        | 29.7      | 40.2   | 2.5      | 291    | 1000.8  | 0.4421               | 4.1634              |
| 2020-06-06 10: 30: 00 | 615                        | 605                        | 31.1      | 34.8   | 3.1      | 306    | 1000.4  | 0.8043               | 5.092               |

GHI: global horizontal irradiation; DNI: direct normal irradiance; RH: relative humidity; AP: atmospheric pressure; WS: wind speed; WD: wind direction;  $\bar{Q}_{10}$ : the mean of 10 consecutive time units;  $\bar{Q}_5^d$ : the mean of 5 consecutive days at the same time; PVOUT: photovoltaic output per unit time.

TABLE 3: Eigenvalue and variance contribution of the input variables.

| Number | Eigenvalue | Variance contribution rate (%) | Cumulative variance contribution rate (%) |
|--------|------------|--------------------------------|---|
| 1      | 6.40       | 23.16                          | 23.16                                     |
| 2      | 5.02       | 20.32                          | 43.48                                     |
| 3      | 4.26       | 17.55                          | 61.03                                     |
| 4      | 3.24       | 10.12                          | 71.15                                     |
| 5      | 2.60       | 7.61                           | 78.76                                     |
| 6      | 1.07       | 4.23                           | 82.99                                     |
| ...    | ...        | ...                            | ...                                       |
| 18     | 0          | 0                              | 100                                       |

**3.3. Evaluation Indexes.** After the completion of the model training, the generation capacity of the PV station in 1 day of continuous time was selected for prediction to judge the accuracy and efficacy of the entire model. Here, the mean square error was used as a loss function to test and measure the model's efficacy. The mean square error was calculated as

$$\text{MSE} = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_{ii})^2, \quad (13)$$

where MSE is the mean square error,  $y_i$  is the actual power generation at time  $I$ ,  $\hat{y}_i$  is the prediction value of the model at time  $I$ , and  $m$  is the amount of sample data. Accordingly, the smaller the mean square error, the higher the model precision.

In addition, the maximum error was used to measure the prediction error range of the model. The larger the value of ME, the worse the accuracy of the model, which was computed as

$$\text{ME} = \max \left\{ \frac{|y_i - \hat{y}_i|}{y_i}, \quad i = 1, 2, 3, \dots, m \right\}. \quad (14)$$

**3.4. Model Building and Experimental Results Analysis.** The experimental environment is Intel i9 processor, Linux + anaconda3 + tensorflow2.0 platform, Spyder software, and Python 3.7 programming language.

The LSTM model structure determines the optimization and prediction accuracy of the training process. In this paper, the LSTM network structure having 4 hidden layers of LSTM and 1 ordinary layer was adopted. The number of neurons in each layer was 512, 256, 128, and 64, respectively.

Specifically, 64 neurons were used as the ordinary layer, and dropout operations were used between layers. The overall structure of LSTM is shown in Figure 3.

Relevant literature discusses the setting of training parameters of LSTM. When the learning rate and attenuation rate are different, the network performance is not the same, and the prediction effect is also different. If LSTM has dropout layers, the probability of dropping neurons is also the key to parameter optimization. According to the comprehensive literature [9, 11], the learning rate is generally small, or the learning rate is degraded dynamically; that is, when the accuracy rate in the network training process is no longer improved, the learning rate will be reduced to one-tenth of the original or lower. The probability of discarded neurons is 0.2 and its vicinity, and the decay rate is 0.9 and its vicinity.

The data taken between 2014 and 2018 were used as the training set, while the data taken from 2019 acted as the test set. After the sample data were standardized, PCA and CCIPCA pretreatment were performed, respectively, and training and testing were conducted on the LSTM. The probability of dropping neurons in the dropout layer was {0.1, 0.2, 0.3}; the attenuation rate was {0.8, 0.9} and the learning rate was 0.001. Moreover, the number of training iterations was 100. The experimental results are shown in Table 4.

Table 4 demonstrates that the prediction accuracy of power generation obtained after CCIPCA processing of the sample data was better than that of PCA + LSTM and LSTM. The LSTM model of  $N = 0.3$  and  $P = 0.9$  was selected to draw the loss function value curve of the training set after 100 iterations. As shown in Figure 4, CCIPCA + LSTM exhibited better convergence as well as a better convergence rate than that of LSTM and PCA + LSTM.

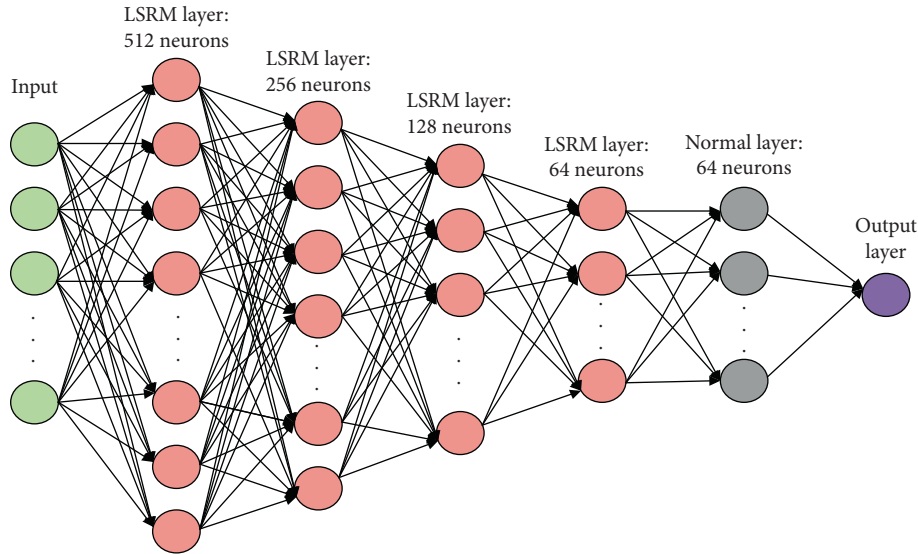


FIGURE 3: LSTM prediction model of the PV station's generation.

TABLE 4: Experimental results of the LSTM model under different parameters of optimization.

| Model         | $P = 0.8$ |           |           | $P = 0.9$ |           |           |
|---------------|-----------|-----------|-----------|-----------|-----------|-----------|
|               | $N = 0.1$ | $N = 0.2$ | $N = 0.3$ | $N = 0.1$ | $N = 0.2$ | $N = 0.3$ |
| LSTM [20]     | 0.081     | 0.072     | 0.064     | 0.079     | 0.078     | 0.074     |
| PCA + LSTM    | 0.079     | 0.070     | 0.058     | 0.073     | 0.062     | 0.061     |
| CCIPCA + LSTM | 0.062     | 0.059     | 0.046     | 0.048     | 0.044     | 0.041     |

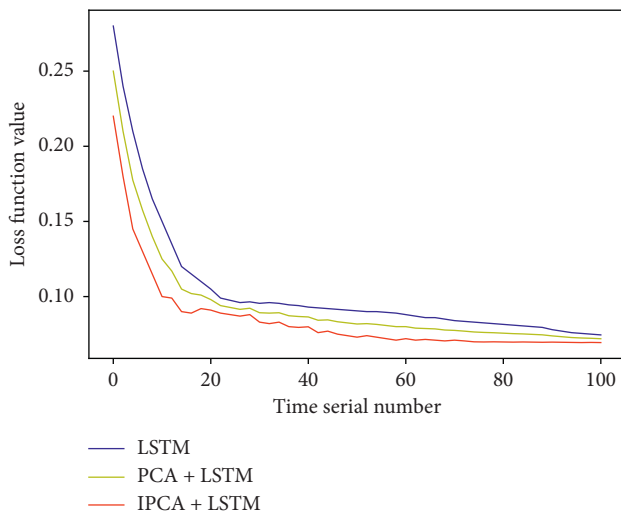


FIGURE 4: Loss function curve of training set and test set.

From the period of 2019, multidays data were selected as test sets in different seasons, and the prediction accuracy of the model was compared with the real power generation.

The data from January 3, 2019, was selected as the test set for comparison with the actual power generation, as shown in Figure 5(a). According to the record, it was light snow, but it was heavy snow three days ago and it turned to clear up after 14:00. On that day, the temperature range was  $-2\text{--}4^\circ\text{C}$ , with no wind, the wind speed range was  $0\text{--}0.4\text{ m/s}$ , and the

absolute humidity was  $1.02\text{--}1.70\text{ g/m}^3$ . Then, the data from January 4, 2019, was selected to verify the model, as shown in Figure 5(b). It was sunny, accompanied by weak solar radiation and breeze, temperature range was  $-5\text{--}1^\circ\text{C}$ , wind speed range was  $0.6\text{--}1\text{ m/s}$ , and the humidity was the same as the previous day.

The meteorological conditions on January 3 and January 4 were similar, and the difference between two days was solar radiation, and solar radiation on January 4 was stronger than that on January 3. Compared with Figures 5(a) and 5(b), the peak of power generation on January 4 was much higher than that on January 3, which indicates that solar radiation is an important factor in power generation prediction. It can also be found from the power generation prediction curve that the photovoltaic power generation efficiency was reduced accompanied by the solar radiation gradually weakening, since the sunlight irradiated the snow surface and reflected, and the ambient temperature sharply dropped in relative to normal times after 14:00 o'clock. When the night came, there was no sunlight and the ambient temperature dropped below zero, and the photovoltaic power generation was zero.

The data on April 3, 2019, were selected as the test set and compared with the real power generation, as shown in Figure 6. According to the data, April 3 was sunny, the temperature range was  $9\text{--}14^\circ\text{C}$ , the wind direction was strong, the wind speed ranged from  $17.2\text{ to }21\text{ m/s}$ , and the absolute humidity was  $2.7\text{--}3.2\text{ g/m}^3$ .

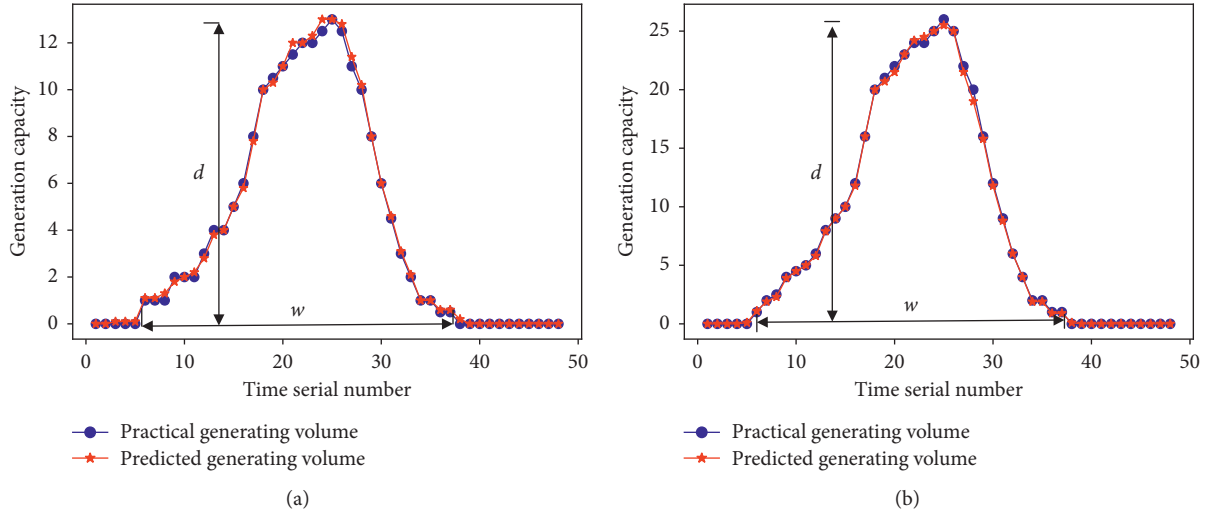


FIGURE 5: Power generation forecast of two days in winter. (a) Comparison between predicted curve and real value of power generation on January 3. (b) Comparison between predicted curve and real value of power generation on January 4. Note:  $w$  represents the time zone of daily effective generation and  $d$  represents the peak value of daily generation.

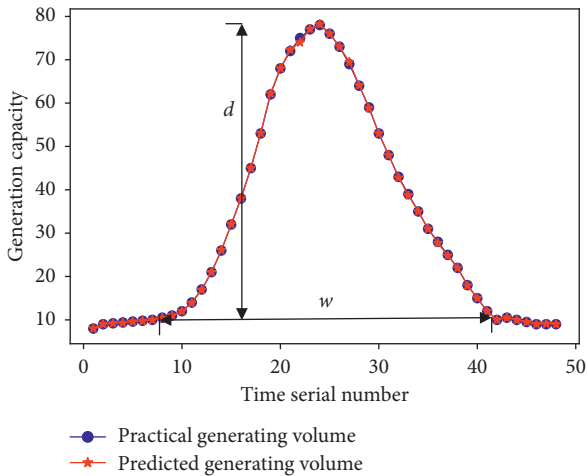


FIGURE 6: Forecast of one-day power generation in spring.

It can be found from Figure 6 that the predicted curve of the whole-day power generation was basically consistent with the real power generation curve, and the two curves were relatively smooth, which indicates that the power generation efficiency of photovoltaic equipment was not affected by wind direction and wind speed.

The data from July 7, 2019, were selected as the test set for comparison with the actual power generation. Figure 7 illustrates that the data recorded exhibited three thunderstorms and severe convection weather on July 7, with periods of 3: 00–4: 00, 7: 00–7: 30, and 14: 00–14: 30, respectively. The meteorological conditions were temperature range of 20–32°C, gust, wind speed of 10.8–13.8 m/s, and absolute humidity of 22.22–25.76 g/m<sup>3</sup>.

Figure 8 illustrates that the photovoltaic power generation efficiency is sensitive to the influence of solar radiation and ambient temperature. In the three periods of July 7, the solar radiation and environmental temperature changed rapidly, and the photovoltaic power generation also changed

rapidly. The prediction results of the model are basically consistent with the real power generation under the condition of rapid changes of photovoltaic power generation factors.

The data on November 11, 2019, were selected as the test set and compared with the real power generation, as shown in Figure 9. According to the record, it was sunny, the temperature range was 9–15°C, the breeze level was 2–3, the wind speed reached 3.1–4.0 m/s, and the absolute humidity was 1.08–2.12 g/m<sup>3</sup>.

From Figures 6 and 7, it can be found that photovoltaic power generation has some relationship with environmental humidity. The meteorological conditions on November 11 and April 3 were basically the same; only the environmental humidity was different, but the peaks of the two pictures were inconsistent, which indicates that the environmental humidity has a negative effect on photovoltaic power generation. On November 11, the environmental humidity was lower, but the power generation effect was better. On April 3, the humidity was higher, and the power generation effect was poor.

It can also be found from Figures 5–7 that the efficiency of photovoltaic power generation is higher in summer,  $d$  in Figure 5 was much higher than that in other graphs, and the photovoltaic power generation efficiency in winter is the worst. Furthermore,  $w$  in Figure 5 was much longer than that in other graphs, which indicates that the photovoltaic equipment uses more time to generate electricity in summer.

The 62-day power generation data from July 1, 2019, to August 31, 2019, were selected as the test set and compared with the real power generation; the prediction accuracy of the model is investigated. As shown in Figure 9, the curve between the actual power generation and the predicted power generation in 62 days was shown.

Figure 9 showed that the two curves were relatively similar, indicating that the prediction accuracy of CCIPCA combined with LSTM was relatively high, and its prediction effect was satisfactory.

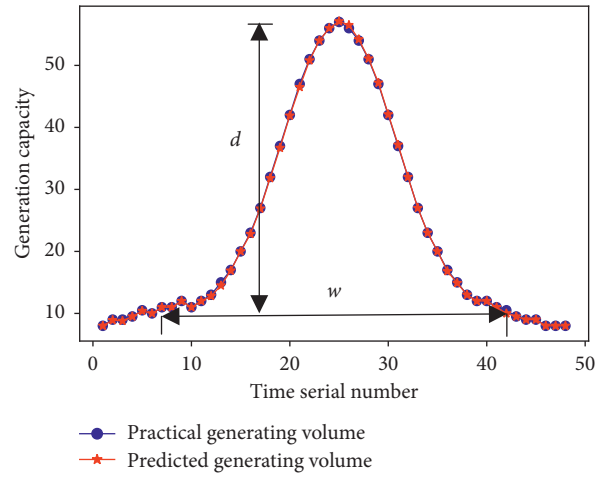


FIGURE 7: Forecast of one-day power generation in autumn.

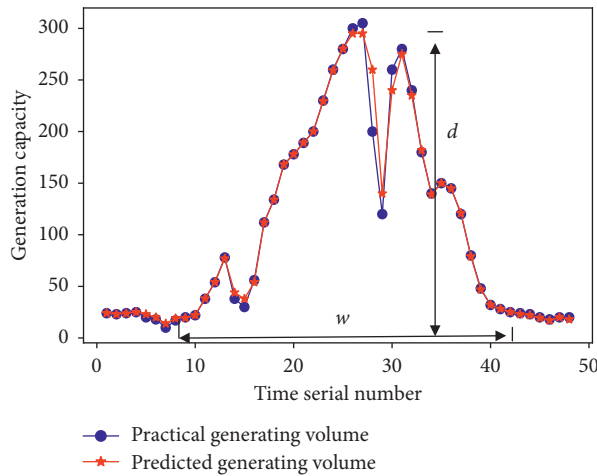


FIGURE 8: One-day power generation forecast in summer.

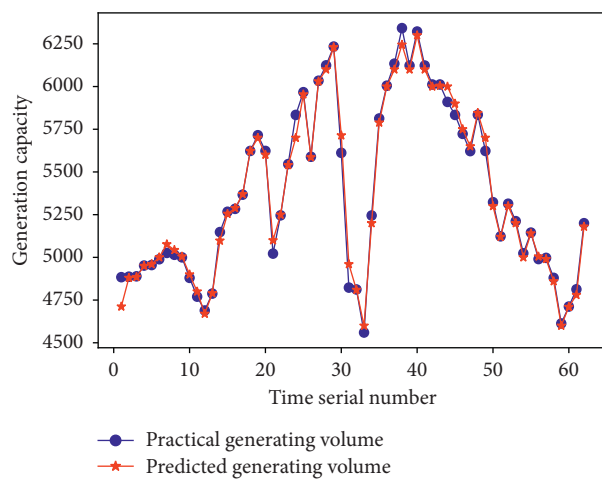


FIGURE 9: Comparison of photovoltaic power generation forecast from July 1 to August 31.

In view of the practicability of the prediction model, horizontal experiments were conducted to compare the prediction model with the GA-BP neural network [21] and Markov chain model (MC) [22], as shown in Table 5.

By comparison, prediction effect of MC is not as good as neural network, while prediction effect of LSTM is higher than GA-BP network due to the advantage of network layer number, and the maximum error is 6.4%; under the



TABLE 5: ME comparison between models.

| Model | ME    | Model       | ME    | Model          | ME    |
|-------|-------|-------------|-------|----------------|-------|
| MC    | 0.074 | PCA + MC    | 0.052 | CCIPCA + MC    | 0.048 |
| GA-BP | 0.068 | PCA + GA-BP | 0.044 | CCIPCA + GA-BP | 0.032 |
| LSTM  | 0.064 | PCA + LSTM  | 0.040 | CCIPCA + LSTM  | 0.023 |

dimension reduction operation of PCA for photovoltaic data, the three prediction models are significantly improved, but, due to the limitations of PCA algorithm, the function mapping from high-dimensional space to low-dimensional space is linear. However, in many practical tasks, it may need nonlinear mapping to find the proper low-dimensional embedding, which leads to the poor dimensionality reduction effect of photovoltaic data. The prediction error of PCA combined with LSTM is 4%, which is higher than that of PCA combined with MC and GA-BP. However, the dimensionality reduction of PV data using IPCA avoids the complex covariance matrix calculation of PCA, and the regularization effect of IPCA is better than that of PCA, which is suitable for LSTM. The prediction error of IPCA combined with LSTM is 2.3%. In a word, the prediction model combining CCIPCA with LSTM displayed better prediction than other models, with an error range of prediction results within 3%.

#### 4. Conclusions

In view of issues such as data being collected by PV stations in a matter of minutes, daily data volume reaching the GB or PB level, data scale being large, multiple instances of data dimensions [23], and existence of noise [24], this paper proposed a prediction model for PV generation by combining CCIPCA with LSTM. According to the simulation of the proposed prediction model, the following conclusions were drawn:

- (1) CCIPCA has handled the super-large-scale data of PV station, realized the dimensionality reduction of data, made use of orthogonalization following dimensionality reduction of data, eliminated the influence of noise, and improved the convergence speed and training speed of the model.
- (2) During model training, the historical horizontal and vertical mean values of PV generation were added to eliminate the disturbance of extreme weather conditions on the model, and 48 sets of data from a certain day were selected for testing. The obtained results aligned with the real values of power generation, demonstrating the model's stable performance.
- (3) The model was compared to the other two PV generation prediction models horizontally, and the power generation prediction error was determined to be less than 3%, illustrating its practicality.

#### Data Availability

The SolarGIS data used to support the findings of this study are included within the article.

#### Additional Points

In the experiment, adopting SolarGIS Meteosat (EUMETSAT, DE) and GOES (NOAA, USA) radiation of satellite remote sensing data, combined with Meteosat (EUMETSAT, DE) and GOES (NOAA, USA) of cloud and snow index and Global Forecast System (GFS) database (NOAA, USA) of water vapor data, a series of meteorological elements including solar radiation and temperature value are calculated. Taking photovoltaic power station no. 11282 in Zhangdian district, Zibo city, Shandong province, China (118 east longitude, 32 north latitude), as an example, the generation data from 2014 to 2019 are selected/studied. All data can be downloaded from <http://solargis.cn>.

#### Disclosure

Some of the authors of this publication are also working on the following related projects: (1) higher vocational education teaching fusion production integration platform construction projects of Jiangsu province under Grant no. 2019 (26), (2) Natural Science Fund of Jiangsu Province under Grant no. BK20131097, (3) "Qin Lan project" teaching team in colleges and universities of Jiangsu province under Grant no. 2017 (15), and (4) high level of Jiangsu province key construction project funding under Grant no. 2017 (17).

#### Conflicts of Interest

The authors declare that there are no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

#### References

- [1] P. Wu, W. Huang, and N. Tai, "Novel grid connection interface for utility-scale PV power plants based on MMC," *The Journal of Engineering*, vol. 2019, no. 16, pp. 2683–2686, 2019.
- [2] A. Ahmad, Y. Jin, C. Zhu, I. Javed, M. Waqar Akram, and N. A. Buttar, "Support vector machine based prediction of photovoltaic module and power station parameters," *International Journal of Green Energy*, vol. 17, no. 3, pp. 219–232, 2020.
- [3] B. Chen and J. Li, "Combined probabilistic forecasting method for PV using an improved Markov chain," *IET Generation, Transmission & Distribution*, vol. 13, no. 19, pp. 4364–4373, 2020.

- [4] P. Christophe, V. Cyril, M. Marc, and M.-L. Nivet, "Forecasting of pre-processed daily solar radiation time series using neural networks," *Solar Energy*, vol. 84, no. 12, pp. 2146–2160, 2010.
- [5] X. Meng, A. Xu, W. Zhao, H. Wang, C. Li, and H. Wang, "A new PV generation power prediction model based on GA-BP neural network with artificial classification of history day," in *Proceedings of the 2018 International Conference on Power System Technology (POWERCON)*, pp. 1012–1017, Guangzhou, China, November 2018.
- [6] G. W. Chang and H.-J. Lu, "Integrating gray data preprocessor and deep belief network for day-ahead PV power output forecast," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 1, pp. 185–194, 2020.
- [7] L. Kaiju, L. Xuefeng, M. Chaoxu, and W. Dan, "Short-term PV prediction based on T-S fuzzy neural network," in *Proceedings of the 2018 33rd Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, pp. 620–624, Nanjing, China, May 2018.
- [8] S. Qijun, L. Fen, Q. Jialin, Z. Jinbin, and C. Zhenghong, "PV prediction based on principal component analysis and support vector machine," in *Proceedings of the 2016 IEEE Innovative Smart Grid Technologies-Asia (ISGT-Asia)*, pp. 815–820, Melbourne, VIC, Australia, December 2016.
- [9] C. Hua, E. X. Zhu, L. Kuang, and D. Pi, "Short-term power prediction of PV station based on long short-term memory-back-propagation," *International Journal of Distributed Sensor Networks*, vol. 15, no. 10, Article ID 155014771988313, 2019.
- [10] E. Oja and J. Karhunen, "On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix," *Journal of Mathematical Analysis and Applications*, vol. 106, no. 1, pp. 69–84, 1985.
- [11] T. D. Sanger, "Optimal unsupervised learning in a single-layer linear feedforward neural network," *Neural Networks*, vol. 2, no. 6, pp. 459–473, 1989.
- [12] J. Y. Weng, Y. L. Zhang, and W. S. Hwang, "Candid covariance-free incremental principal component analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 1034–1040, 2003.
- [13] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, vol. 24, no. 6, pp. 417–441, 1933.
- [14] L. Yao and Y. Guan, "An improved LSTM structure for natural language processing," in *Proceedings of the 2018, IEEE International Conference of Safety Produce Informatization (IICSPI)*, pp. 565–569, Chongqing, China, December 2018.
- [15] H.-X. Tian, D.-X. Ren, and K. Li, "A hybrid vibration signal prediction model using autocorrelation local characteristic-scale decomposition and improved long short term memory," *IEEE Access*, vol. 7, pp. 60995–61007, 2019.
- [16] B. Zhang, S. H. Zhang, and W. H. Li, "Bearing performance degradation assessment using long short-term memory recurrent network," *Computers in Industry*, vol. 106, pp. 14–29, 2018.
- [17] M. Li, D. Yu, Z. Chen, K. Xiahou, T. Ji, and Q. H. Wu, "A data-driven residual-based method for fault diagnosis and isolation in wind turbines," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 2, pp. 895–904, 2019.
- [18] G. J. Liu, H. X. Gu, X. C. Shen, and D. You, "Bayesian long short-term memory model for fault early warning of nuclear power turbine," *IEEE Access*, vol. 8, pp. 50801–50813, 2020.
- [19] A. H. Mirza, M. Kerpici, and S. S. Kozat, "Efficient online learning with improved LSTM neural networks," *Digital Signal Processing*, vol. 102, Article ID 102742, 2020.
- [20] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841–851, 2019.
- [21] C. Dai, J. Wu, D. Pi et al., "Brain EEG time series clustering using maximum weight clique," *IEEE Transactions on Cybernetics*, vol. 99, pp. 1–15.
- [22] L. Cui, J. Wu, D. Pi, P. Zhang, and P. Kennedy, "Dual implicit mining-based latent friend recommendation," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 5, pp. 1663–1678, 2020.
- [23] C. Dai, D. Pi, S. I. Becker, J. Wu, L. Cui, and B. Johnson, "CenEEGs," *ACM Transactions on Knowledge Discovery from Data*, vol. 14, no. 2, pp. 1–25, 2020.
- [24] W. Shao, D. Pi, and Z. Shao, "A pareto-based estimation of distribution algorithm for solving multiobjective distributed No-wait flow-shop scheduling problem with sequence-dependent setup time," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 3, pp. 1344–1360, 2019.

## Research Article

# Analysis and Simulation of Multimedia English Auxiliary Handle Based on Decision Tree Algorithm

**Kaiwei Yan** 

*School of Foreign Languages, China West Normal University, Sichuan, Nanchong 637002, China*

Correspondence should be addressed to Kaiwei Yan; [tracey@cwnu.edu.cn](mailto:tracey@cwnu.edu.cn)

Received 14 July 2020; Revised 26 August 2020; Accepted 28 August 2020; Published 15 September 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Kaiwei Yan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, through the improved decision tree algorithm, the handles in multimedia English assistance are parsed and simulated. In order to better perceive the sense of language in English composition and improve the rationality of intelligent evaluation, an N element based on association analysis is proposed. Sense value quantification calculates its support in the corpus by obtaining N-tuples of the composition. If the degree of support is lower than the threshold, the part where the language sense problem occurs is analyzed, and the type of language sense problem is judged for the students to provide assistance in modifying the composition. In addition, this paper also extracts word features, sentence features, and text structure features in the composition to fit the English handles analytical score. By testing the test set, the experiment shows that, by extracting the language sense features of the candidate's English composition, it can not only judge whether there is a problem with the language sense of the candidate, but also provide a basis for the overall evaluation of the composition.

## 1. Introduction

With the rapid development of information technology, people can use mobile phones, handheld computers, and other handheld mobile devices to obtain, process, and send information at any time or place, so that communication is everywhere, information is everywhere, and we also rely on handheld mobile devices. Carrying out educational activities and transmitting educational information with wireless networks has provided the possibility for human lifelong learning [1]. In this era of the popularity of handheld mobile devices, especially smart phones and tablets, almost every college student has one. In this environment, where learning can be done anytime, anywhere, people's learning habits and behaviors are quietly changing, and fragmented learning has also become one of our main learning methods. The learner's pronunciation problem is influenced by the phonological system mainly in that pronunciation-related organs or pronunciation actions are not standard, and there are large differences in phoneme discrimination. Learners often pronounce by blind imitation. They cannot fundamentally recognize how

to pronounce correctly, and pronunciation problems cannot be found in time, and if they are not found, there is no feedback to correct them. Therefore, learners sometimes do not even know whether their pronunciation is standard. In order to ensure that your pronunciation is correct, there are many people who are willing to pay high tuition fees and ask foreign teachers to correct their pronunciation. With the popularity of online language learning, an automatic pronunciation error detection and correction system has been spawned [2]. At present, there are only a few products on the market that has pronunciation problems, and most of the functions are relatively simple. Learners can only imitate the audio and video learning materials played by follow-up, and then the system plays the recording. There are only a few software that has feedback for detecting spoken pronunciation, but the flaw is that the feedback function is not enough to solve the root problem of the learner [3]. This function is only available after the learner has followed the pronunciation. It can be pointed out that the learner's pronunciation is not good enough, but the learner cannot understand where his pronunciation is wrong, and how to improve the

pronunciation, so that the learner cannot get the most valuable feedback to correct the information, and often this does not improve the learner's oral ability [4, 5].

The current research related to providing guidance for logging statement-level decisions is not that rich. In the empirical research on long practice, Ji et al. found that the revision history of logging statements by developers accounted for 72% of the changes at the logging statement level, so a simple logging statement level check was designed [6]. Their checker principle is based on the observation if the log code in a similar code block has an inconsistent logging statement level. Pan infers part of the execution path by mapping log messages to source code [7]. Cai et al. analyzed the logs to understand the correct dependencies between log messages from normal execution and used this information to identify anomalies in failed execution [8]. In addition, machine learning and data mining technologies also show great potential in tracking the scale and complexity of large-scale system monitoring and diagnostic challenges [9, 10]. Some studies learn statistical features to detect and diagnose abnormalities. Lai and Chen used classification techniques to group similar log sequences into a set of classes based on certain string distance metrics [11].

Subecz Z first extracted  $n$  graphs as the features of the system call sequence and then used support vector machines to classify the trajectories according to the similarity of the trajectories of known problems [12]. Ma et al. introduced the specific implementation of the text classification system in detail, which provided a basis for the subsequent text classification research [13]. McLarnon and O'Neill produced the L1-L2MAP tool, which included manually input phoneme data and then used this data to create a list of expected pronunciation errors [14]. Similarly, for learners whose mother tongue is Vietnamese, Professor Ha's research team studied the common phoneme substitution errors between Vietnamese and English [15]. Yan studied the distinguishing characteristics of flat tongue and tongue-twisted sounds. The research results show that there is a big difference between the peaks of the spectrum energy. Therefore, the energy concentration segment is selected as the distinguishing feature to discriminate between flat tongue and tongue phonemes [16]. Ma and Chow studied the pronunciation of consonants by Japanese learners while learning English. The study shows that, due to the aspirated variants of unvoiced stop sounds in Japanese, the use of non-aspirated/aspirated feature as a distinctive pronunciation feature can well distinguish the categories of consonant pronunciation [17]. An improvement of 17.35% reduction in error rate can be achieved. Finally, they also applied this technology to the CAPT system [18].

At present, although there are many related researches on the location of logging statements, they provide corresponding guidance for developers in the industry, but the relevant content in the recording content is relatively insufficient. In order to fill the gap, it provides developers with practical work. Logging statement-level recommendation functions. This paper implements a text classification technology based on machine learning and recommends levels through code block-related features. The paper conducts a comprehensive research on text classification technology

based on machine learning. The process and the text reprocessing, text representation, spatial dimensionality reduction, classification method, and classification performance evaluation were analyzed and discussed. After that, we focus on the random forest algorithm in the classification algorithm and propose improvements based on the analysis of its principles and characteristics [19]. The improvement mainly includes two aspects: on one hand, the paper explores and optimizes the handle simulation mechanism of the decision tree classification algorithm, performs weighted handle simulation based on the classification effect and prediction probability of the decision tree, and uses the weighted handle simulation to improve the traditional random forest classification algorithm. In the handle simulation mechanism, and text classification experiments to verify the improved effect. On the other hand, the concept of hyperparameters in the algorithm and commonly used hyperparameter adjustment algorithms are introduced. Based on the analysis of text classification scenarios and the random forest algorithm, an algorithm based on random handle simulation and grid handle simulation is proposed to carry out the random forest algorithm. Parameter optimization and, finally, design experiments prove the effectiveness of the algorithm.

## 2. Improved Decision Tree Algorithm for Simulation Design of English Handle Resolution

*2.1. Improved Decision Tree Algorithm.* Decision tree algorithm is an instance-based inductive learning algorithm. It focuses on inferring the classification rules in a set of unordered and irregular examples, which are represented by decision trees. The purpose of constructing a decision tree is to get the relationship between attributes and categories and use this relationship to predict the category labels of samples of unknown categories. The algorithm uses a top-down recursive method, compares the attributes of the nodes within the decision tree, judges the branch of the node based on different attribute values, and draws conclusions from the leaf nodes of the decision tree. The main decision tree algorithms are ID3, C4.5 (C5.0), CART, PUBLIC, SLIQ, and SPRINT algorithms. The decision tree classification method has the advantages of processing both data-type and regular-type attributes, insensitivity to intermediate values, processing samples with missing attribute values, and easy to understand output results [20]. However, there are shortcomings that are prone to overfitting, and the results are biased toward features with more values, as shown in Figure 1. The decision tree algorithm is a meta-classifier of the random forest algorithm studied in this paper, which will be studied in more depth below.

There are various evaluation indicators according to different text classification application backgrounds. The metrics for classification performance evaluation include error rate, accuracy rate, recall rate, accuracy rate,  $F$  balance, microaverage and macroaverage, and ROC curve. The error rate is the proportion of samples with incorrect classification results to the total number of samples, and the accuracy rate is the proportion of samples with correct classification results to

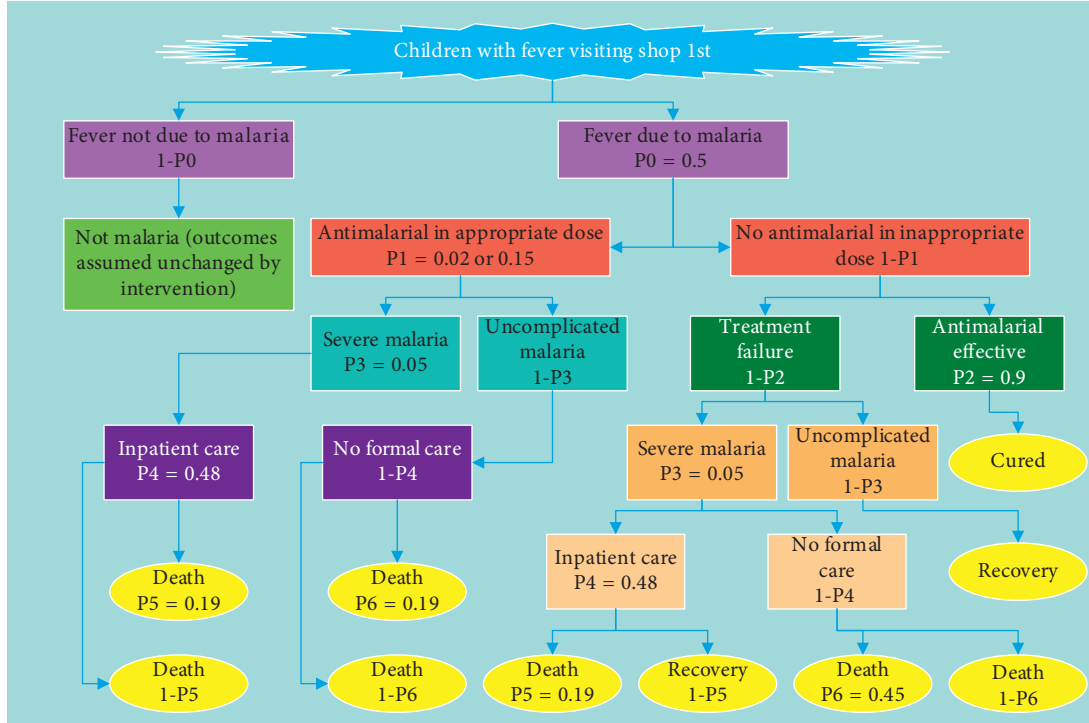


FIGURE 1: Improved decision tree algorithm model.

the total number of samples. They can be used for binary classification or multicategory. But this does not comprehensively reflect the performance of classification models, especially the problem of unbalanced classification [21]. The recall rate and accuracy rate can be calculated based on the confusion matrix established by the classification results. The confusion matrix is shown in Table 1. TP indicates that the text in category C is judged to belong to the text of category C, that is, the text with the correct classification result in category C. FP Text that does not belong to category C is judged to belong to category C; that is, text that is not classified in category C is incorrect, FN is the classification error in category C, and TN is the classification result in noncategory C. The recall rate is the ratio of the number of samples correctly judged to be in this category and the total number of samples belonging to this category, also called the recall (Recall), and the calculation method is formula (1).

$$\text{STFT}[A(t, w)] = \int_{-\infty}^{+\infty} a(\tau)g(\tau - t)e^{-jw\tau} d\tau. \quad (1)$$

Precision is the proportion of samples whose true category is C in the number of samples whose classification result is C, also known as the precision rate, and the calculation method is

$$R_m = \frac{E}{A \times F} * \beta. \quad (2)$$

$F$  equilibrium is an evaluation index that takes into account both the recall rate and the accuracy rate, and its calculation method is shown in formula (3), where  $\beta$  is a value greater than 0, as well as the accuracy and recall weight adjustment parameters.  $\beta < 1$  has a greater impact on the

accuracy rate, and  $\beta > 1$  has a greater impact on the recall rate. When  $\beta = 1$ , the recall rate and accuracy rate have the same weight, and the  $F$  equilibrium at this time is the  $F1$  indicator.

$$R_f = \frac{c}{A \times D} * \beta. \quad (3)$$

Microaverage and macroaverage are the evaluation indicators when evaluating the classifier on the entire data set. They average the results of all categories. They are different from the single-category classification evaluation indicators such as recall rate, accuracy rate, and  $F$  balance. The calculation methods of microaverage and macroaverage are formula (4) and formula (5), respectively. Micro-averaging first calculates the total number of instances with correct and incorrect classification results in all categories and then calculates the recall rate and accuracy rate of all categories. It also emphasizes the impact of the category with a larger number of samples on the overall result. Macroaverage first calculates the recall rate and accuracy rate of each category and then performs arithmetic average to get the recall rate and accuracy rate of all categories.

$$GJ_{ik} = \frac{GJ_{ik}^{jint}}{\sum_{i=1}^d GJ_{ik}^{jint}}, \quad (4)$$

$$D = \frac{(n-1)d_n}{\sum_{i=1}^{n-1} d_i}. \quad (5)$$

The full name of the ROC curve is the Receiver Operating Characteristic Curve, which is a curve in a two-dimensional plane. The horizontal coordinate of the plane is

TABLE 1: Classification result confusion matrix.

| Type                                    | Text for category C | Noncategory C text |
|---|---------------------|--------------------|
| Judge text that belongs to category C   | 14                  | 22                 |
| Text that does not belong to category C | 18                  | 42                 |

negative positive rate (False Positive Rate, FPR), which represents that the sample predicted by the classifier as positive is actually negative. The proportion of samples in all negative samples is True Positive Rate (TPR), which represents the proportion of actual positive samples to all positive samples in the samples predicted by the classifier as positive [22]. The ordinate is True Positive Rate (TPR), which represents the proportion of actual positive samples to all positive samples in the samples predicted by the classifier as positive. The curve is usually located on the coordinate lines (0, 0) and (1, 1), the area under the ROC curve is AUC, it is a number, and the effect of the classifier can be evaluated more intuitively. The larger the value, the better the classification effect.

The decision tree algorithm is the base classifier of the random forest. It is an inductive learning algorithm. It learns the classification rules in the form of a tree structure with a large number of samples without order and rules and uses this rule to predict unknown samples. The decision tree is composed of nodes and directed edges, and nodes have intermediate nodes and leaf nodes. Each intermediate node has 4 parameters. One is the decision function, which is the value of a feature. When the feature is less than or equal to this value, the decision path is to the left, and when the feature is greater than this value. The decision path is to the right. The second is the impurity value, which reflects the prediction ability of the current node. The third is the number of covered samples, which refer to the number of samples participating in the decision of this node. The more the covered samples, the more stable the decision functions.

The learning process of decision tree mainly includes feature selection, decision tree generation, and decision tree pruning. Feature selection is to select features from the features of the training data as the segmentation criteria of the current node, where different criteria for feature selection produce different decision tree algorithms. Decision tree generation is a recursive process that generates child nodes recursively from top to bottom according to the selected feature evaluation criteria and stops generating after reaching certain constraints. Pruning is because decision trees are prone to overfitting. To reduce overfitting of decision trees, methods such as prepruning and postpruning can be used to reduce the structure and size of the decision tree. The core algorithm for decision tree growth is the branching criterion, including how to select the best grouping variable from many variables, and how to select the best segmentation point of multiple values of the grouping variable. According to the different methods of decision tree node splitting, it can be divided into two categories; one is ID3, C4.5, and other decision trees based on information entropy splitting, and the other is the CART decision tree based on the Gini index classification.

The full name of CART is Classification and Regression Trees, which can do both classification and regression. It is a classification tree when the result to be predicted is a discrete value, and a regression tree when the result to be predicted is a continuous value. CART is a binary tree, and non-leaf nodes have two branches because it recursively divides the samples on the current node into two subsets during the node splitting process. CART uses the Gini coefficient when the nodes are split and uses the Gini coefficient minimization principle. The calculation method of the Gini coefficient is

$$k_{nm}(x) = \sum_{i=1}^N w_i s_i(x) = W^T S(x), \quad (6)$$

$$M^* = \arg \min \left\{ \sup_{x \in \Omega_x} |b(x) - W^T S(x)| \right\}. \quad (7)$$

Constraints for the decision tree to stop growing, that is, to no longer perform node splitting, include the node's impurity purity reaching the threshold, the node's sample number reaching the threshold, the attributes to be split reduced to a certain value, and the depth of the decision tree growing to a certain value.

*2.2. Design of Multimedia English Auxiliary Handle Analysis Simulation System.* The main function of this system is to provide a self-learning method, especially for those who have left school but want to take the exam. It is difficult to get the teacher's timely examination. Students may not approve test questions, and teachers may not have time to approve a large number of test papers. This system is a learning system for simulated exams developed to solve the actual needs of this society. The overall design of this system is designed based on the independence of the modules. Module independence is in the design and development process, the function of each module is independent, and it has less interaction with other modules. The independence of modules can make effective modular software easy to develop, and independent modules are more convenient for testing and maintenance. The modules of this system are loosely coupled with each other, and the internal elements of the modules are closely combined.

The system contains a module of question bank management and test paper management. The question bank management module provides teachers with the functions of adding, modifying, and deleting test questions. The test paper management module adds or deletes the test questions in the question bank to the test paper. When adding test questions, it is only the question bank management module that has an impact. Since the test paper just added does not

add to the test paper, it does not have any impact on the test paper management module. When adding test questions in the test paper, only rely on the foreign key to generate the association relationship, so as to form a test paper. This shows that the independence between the two modules is high, so the coupling relationship is loose, as shown in Figure 2.

For the test bank management module, its function is to maintain the test bank. You can add, delete, modify, and check the test questions. You can only operate the large and small question databases in the database, and you cannot and are not allowed to operate the content in other databases. That is to say, the only thing the question bank management module does is to maintain the test question bank. Things that are not within its responsibilities can only be requested by other modules to complete. This is called high cohesion.

In addition to the basic requirements of the system, there are the most important functional business requirements for pronunciation correction. Automatic pronunciation correction system mainly includes two functional modules: a user information management module and a pronunciation correction module. The user management module includes student users, teacher users, and system administrators. Pronunciation correction module includes pronunciation data collection module and pronunciation data error detection module, pronunciation data correction module, and historical data display module. Figure 3 is a functional block diagram of the automatic pronunciation error correction system.

The server is the core of the automatic pronunciation correction system. All functions are implemented through the server Java code. The server uses the framework of spring + spring MVC + Mybatis. Among them, spring is like a container to automatically create and manage the instances in the project. By configuring parameters in the configuration file to call Java entity classes to create objects, its core idea is control inversion (IOC) and dependency injection (DI); using the Spring framework to develop projects can enable developers to focus on the development of business functions and improve development efficiency without paying attention to the creation and management of instance objects. Spring MVC based on Java is a popular lightweight Web MVC pattern framework, which simplifies our development through a request-response driving model. MyBatis is a database interaction layer framework used by many companies. MyBatis does not need JDBC code to operate the database and result set. Simply configure the mapping relationship in XML to complete the mapping between database fields and Java entity class attributes.

*2.3. Analysis and Simulation Evaluation Analysis.* The running efficiency of the algorithm can be evaluated by the time and space resources required by the computer when running the algorithm. The time resource required by the computer when running the algorithm is also called the time complexity of the algorithm, which mainly

depends on the following factors: the time required for the input of the algorithm data, the time required for the algorithm to be compiled into an executable program, and the time required for the computer to execute each instruction, and the number of times to execute the algorithm statement repeatedly. Since the first three factors mainly depend on the performance of the device, it is customary to use the number of times the algorithm statement is repeatedly executed as the time complexity of the algorithm. When comparing, generally do not care about its precise measurement; only care about the order of magnitude. In the actual environment, there is a more convenient way to measure the time complexity of the algorithm, for example, by comparing the time difference before and after the execution of the algorithm on the same machine.

The storage space resource occupied on the computer memory when running the algorithm is also called the space complexity of the algorithm. The space complexity of the algorithm includes static storage space, storage space required for the input and output data of the algorithm, and variable storage space. The static storage space is a fixed part, and its size does not have a great relationship with the needs of users, including the space occupied by the algorithm code and the space occupied by variable constant characters in the algorithm. The storage space required for the input and output data of the algorithm depends on the problem to be solved. The input and output data are passed through the calling function, and the storage space they need is not different due to the difference of the algorithm. Variable storage space refers to some auxiliary space temporarily generated when running the algorithm. This part has no great relationship with the algorithm itself but is related to the auxiliary algorithm storage structure called by the operating system to execute the algorithm. These spaces are larger than those required by the algorithm itself and are issues that require objective consideration. Because of the rapid development of hardware technology, computers have more and more storage space, and the limitation of storage capacity has little effect on the algorithm. When analyzing the running efficiency of the algorithm, we cannot consider the space complexity of the algorithm more and consider the time complexity of the algorithm more, as shown in Figure 4.

The value range of Brier Score is 0 to 1. The lower the score, the better the performance of the model. Brier Score represents the error between the predicted probability and the actual observation value. That is, the probability of the predicted logging statement level is the same as the actual logging statement level. Probability prediction can assign a very high probability correct logging statement level to instance data or just assign the correct category with a probability that is only slightly higher than the error level probability to instance data. Brier Score is conducive to identifying the ability of the classification model to accurately predict the corresponding category in the former case and can also explain the performance of the classifier model in predicting the category compared to random guessing.

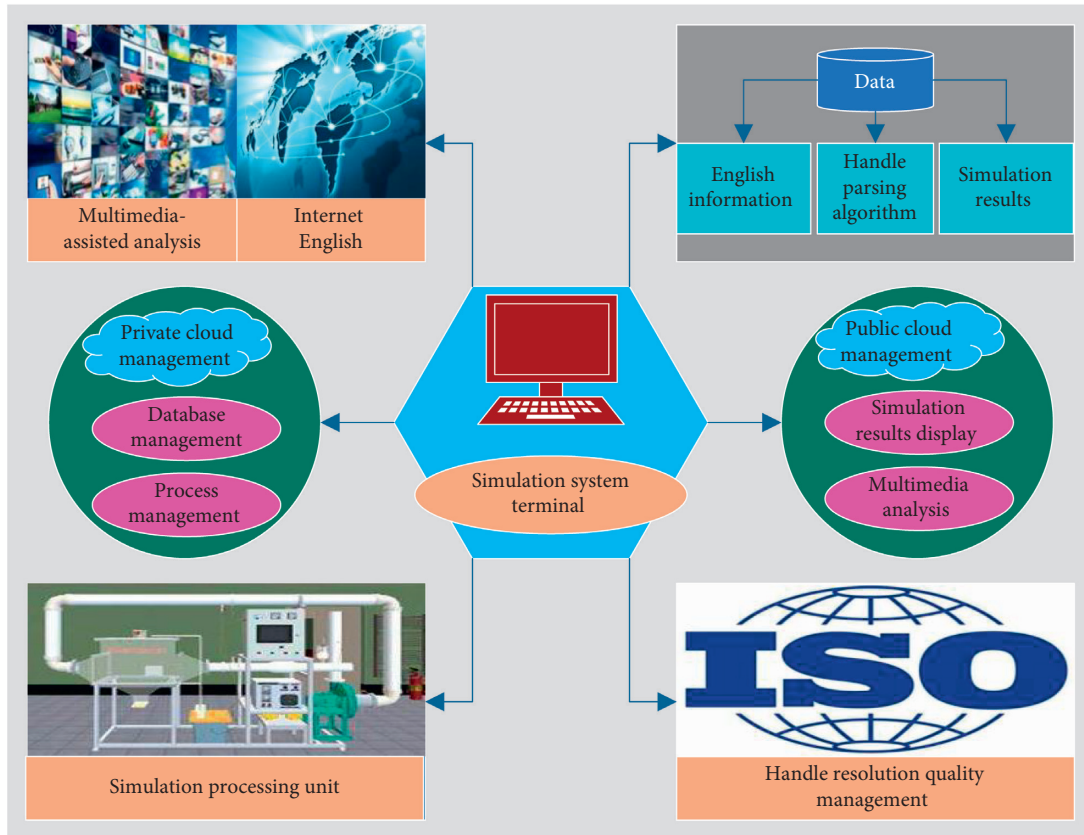


FIGURE 2: Multimedia English auxiliary handle parsing simulation system design.

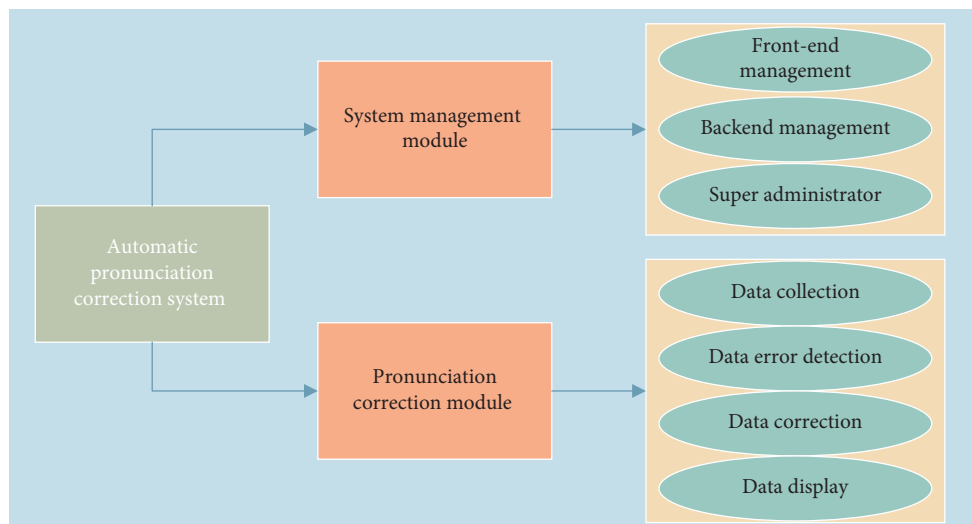


FIGURE 3: Functional block diagram of automatic pronunciation correction system.

### 3. Result Analysis

3.1. *Evaluation of Experimental Results.* The main work of this paper is to use machine learning text classification technology to implement a logging sentence level recommendation method. As an important source of experience for machine learning, this article selects the top 100 ranked GitHub and uses the logging statements in the Java

language project to build a classification model. The AUC and Brier Score performance results of the classification models constructed by the three classic classification algorithms are shown in Figure 5.

It can be seen from Figure 5 that the AUC of the three classification algorithms is up to 0.815, and only 0.798 when the learning effect is poor. Although the experimental dataset in this paper (25 GitHub projects) is



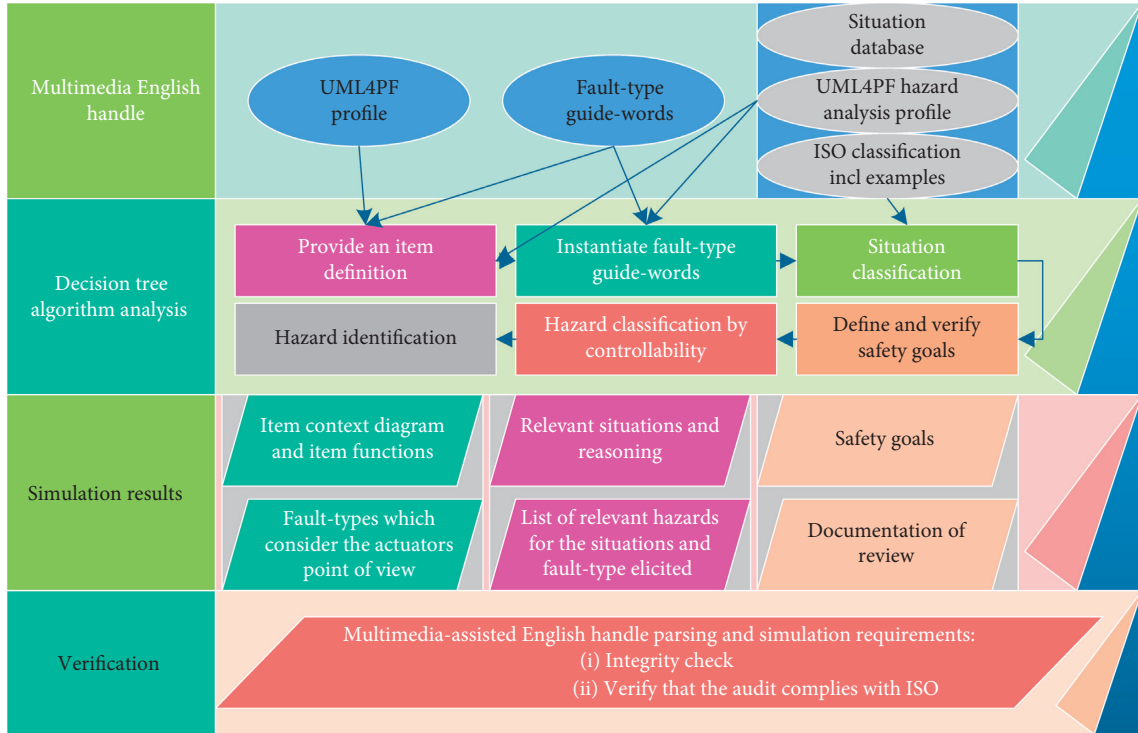


FIGURE 4: Analysis and simulation evaluation design.

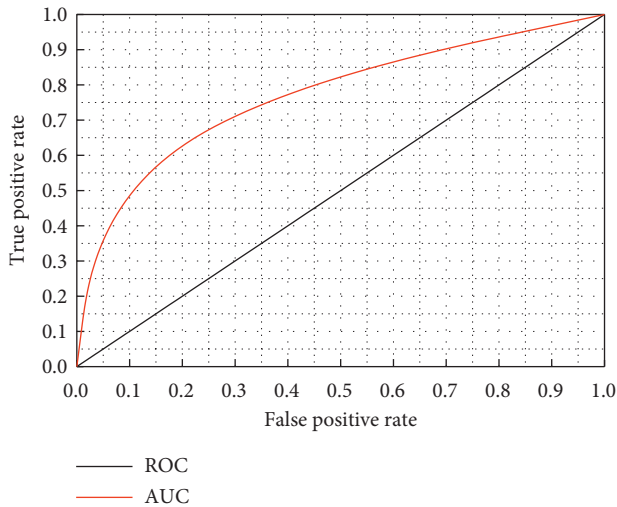


FIGURE 5: ROC and AUC after decision tree classification.

different from the approximate study in the field of logging statement-level recommendations (4 open source projects), it is different from the AUC score in the approximate study (0.75 to 0.81). In contrast, the experimental results of this article are generally equivalent to them, with a slight advantage. Moreover, considering that the increase in the number of items may lead to an increase in the difference in the characteristics of logging statements, the AUC score of 0.798 to 0.815 is sufficient to illustrate that the text classification model constructed in this paper can well overcome the uneven distribution of cross-project machine learning predictions. And

regarding other challenges, it has a better performance in distinguishing different levels of logging statements.

On the other evaluation standard, Brier Score, three classification algorithms selected in this paper can reach 0.440 at the best time and 0.462 at the worst time. Compared with the Brier Score of 0.44 to 0.66 in the approximate study, it has better performance. It shows that, as a probabilistic prediction algorithm, the machine learning model constructed in this paper has a high accuracy rate, which provides a reliable guarantee for the correct recommendation of logging statement level.

It can be seen from the two evaluation indicators that the decision tree is the best among the three selected algorithms. In the ROC curve, one point that needs attention is (0, 1), which represents  $FPR=0$  and  $TPR=1$ , which means that  $FN=0$  and  $FP=0$ , and the classifier in this state can use all the samples. The datum is classified into the correct category, which means that it is an ideal classifier. Therefore, the closer the value of AUC to 1, the stronger the classification performance of this classifier.

This article conducted multiple samplings and then split the sample data in the same way according to the ratio of 9:1. 90% of the data was used for training, and the remaining data was used as test data to verify the performance of the model. Considering the influence of the number of data on the performance of the classification model, this paper selects from four to four representative sampling data of the experimental results for display, as shown in Figure 6, and the number of logging statements for each sample data has been Callout.

It can be seen from Figure 6 that when the number of items decreases and the differences in the factors that affect

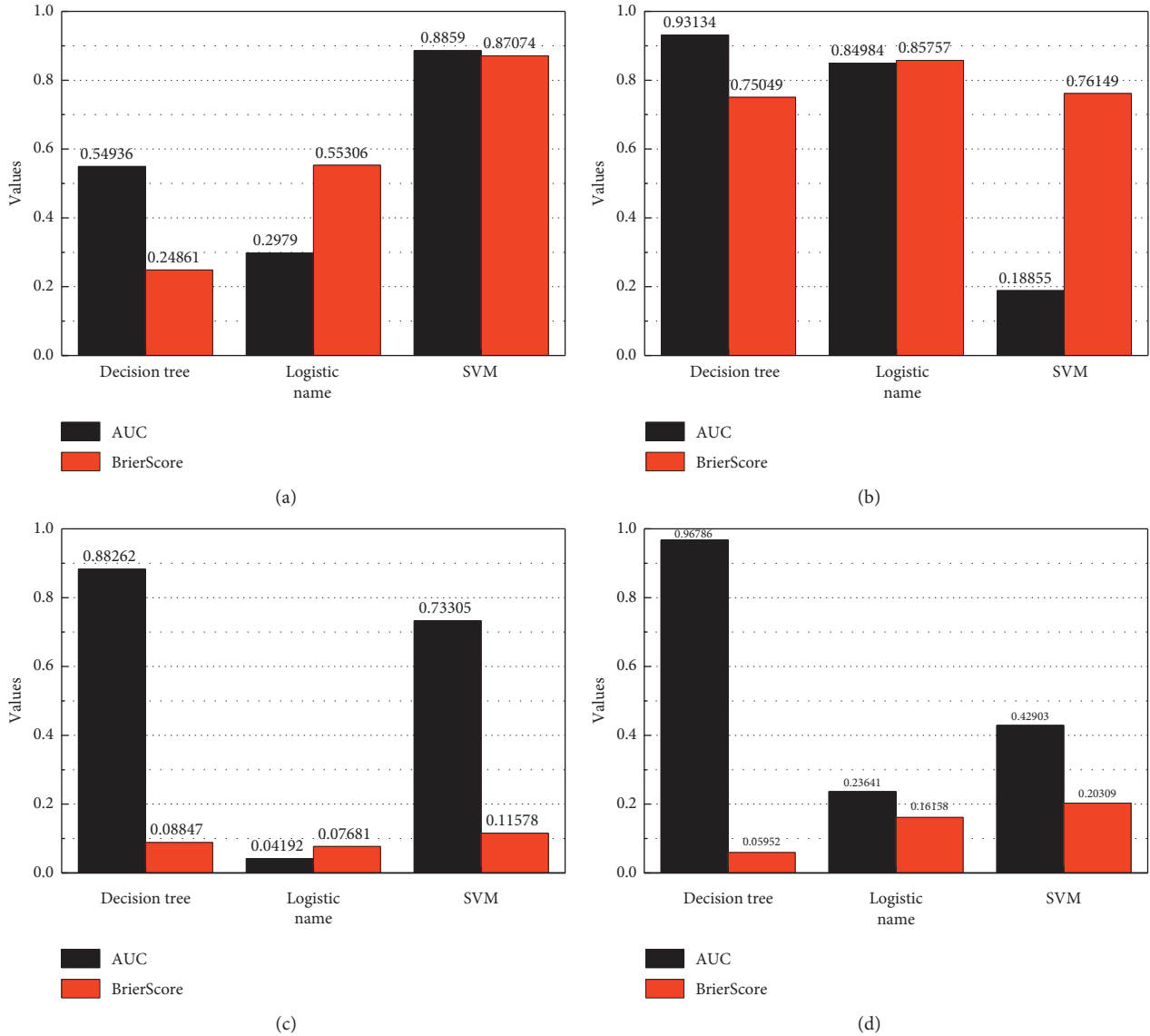


FIGURE 6: Sample experimental results.

the logging statement level become less, the method proposed in this article performs excellent in both AUC and Brier Score indicators. Compared with the test results on multiple projects, the AUC of each sample experiment has a certain degree of improvement, and it can reach 0.836 when the performance is the best. Such a high value means the distance between its ROC curve and the upper left corner. In the process of classifying the log statements, it is likely to be classified into the correct category, and the prediction results of the model are very reliable. The Brier Score indicator has also made no small progress, from the lowest 0.440 on the multi-project data set to the lowest 0.373, successfully below 0.4, and closer to 0, indicating that there is feature learning on approximate regular data; the method proposed in this paper can accurately predict the level of logging statements, and the error in probability prediction is small.

*3.2. Optimal Decision Tree Algorithm Handles Analysis Mechanism Result Analysis.* Aiming at the traditional random forest classification algorithm and the random forest classification algorithm optimized in this paper to optimize the handle simulation mechanism in text classification, a comparative experiment is conducted. The experimental data use all the data in the 20 news by date data sets, use the training set in the data set to train the classifier, and use its test set to evaluate the classification effect of the algorithm, and the selection time of the evaluation indexes of the performance of the two classification algorithms and the accuracy rate of the algorithm prediction. The time is the time used by the algorithm to train the data set to obtain the classifier and use the classifier to predict the test set for the category result. The prediction accuracy of the algorithm is the ratio of the number of samples that the algorithm

predicts the test samples to obtain the correct classification result to the total number of test samples. Since the number of decision trees in the random forest will have an impact on the performance of the algorithm, when conducting a comparative experiment on text classification using the traditional random forest classification algorithm and the random forest classification algorithm with the optimized handle simulation mechanism in this paper, we should ensure that the number of decision trees is excluded. Under the same conditions except for others, choose different numbers of decision trees for text classification. In the experiment, choose the number of decision trees as 10, 30, 50, 100, 200, 300, 400, and other hyperparameters are the default values for comparison experiments. In addition, in order to eliminate the effect of randomness on the experimental results, comparison experiments under the same number of decision trees are conducted 10 times, and then the average value of the 10 results is taken as the final result.

According to the experimental design, the text classification experiment is carried out, and it is concluded that, under different numbers of decision trees, the traditional random forest classification algorithm and the random forest classification algorithm with optimized handle simulation mechanism in this paper are used. The time for text classification is shown in Figure 7.

It can be seen from Figure 7 that the number of decision trees in the random forest classification algorithm has a certain impact on the time of text classification under the two algorithms. As the number of decision trees in the random forest classification algorithm increases, the two algorithms are used for text classification. The time spent is increasing accordingly. Figure 7 lists the time to use two algorithms for text classification under different numbers of decision trees. Although the random forest classification algorithm that optimizes the handle simulation mechanism in this paper is more complicated than the traditional random forest algorithm used in handle simulation, according to the prediction training accuracy of the out-of-bag data, each decision tree is given a certain weight. When the handle simulation is performed, it is not just outputting the category of the sample, but the probability that the sample belongs to each category. There is no significant increase.

The traditional random forest classification algorithm and the random forest classification algorithm with optimized handle simulation mechanism in this paper are used for text classification experiments. The prediction accuracy of the two algorithms is shown in Figure 8.

It can be seen that the number of decision trees in the algorithm has a greater impact on the accuracy of the algorithm. Generally speaking, the larger the number of decision trees, the higher the accuracy of the algorithm prediction. In the case of different decision tree numbers, the prediction accuracy of the random forest classification algorithm using the optimized handle simulation mechanism in this paper is improved to a certain extent compared with the traditional random forest algorithm. Combining the time comparison used in the text classification experiment using the two algorithms in Figure 8, the random forest

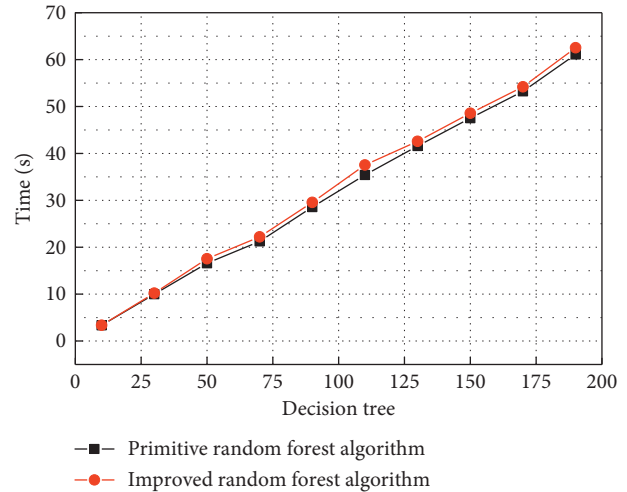


FIGURE 7: Algorithm time comparison.

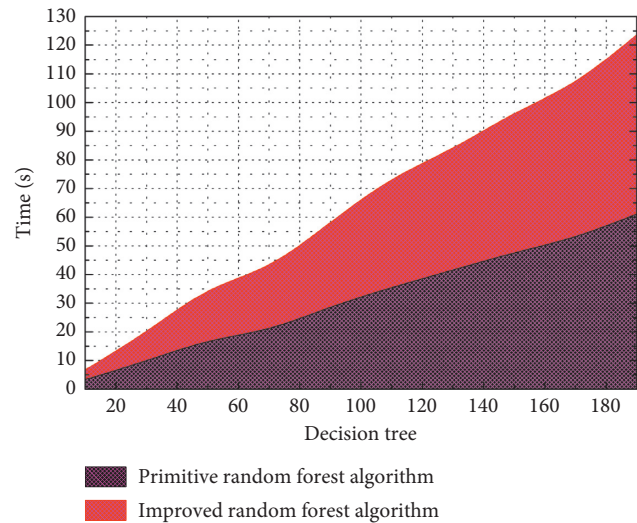


FIGURE 8: Comparison of algorithm prediction accuracy.

classification algorithm that optimizes the handle simulation mechanism in this paper improves the prediction accuracy of text classification on the basis of almost no increase in time resources, compared with the traditional Random forest algorithm that has higher performance. In addition, the prediction accuracy of the two algorithms has not reached a high level, which is related to the use of default values for the hyperparameters of the algorithm except for the number of decision trees, which also illustrates the importance of optimizing the hyperparameter values of the algorithm.

**3.3. Handle Simulation Result Analyses.** The design comparison experiment proves that combining the optimized handle simulation mechanism and the proposed hyperparameter optimization algorithm has improved the effectiveness of the traditional random forest algorithm. Because the values of the number of decision trees and the number of feature attribute subsets in the random forest algorithm have

a great influence on the performance of the random forest algorithm, these two hyperparameters are selected in the experiment. First, the proposed hyperparameter optimization algorithm is used to find the hyperparameter optimization value of the random forest algorithm with the optimized handle simulation mechanism, then two text classifications are performed, and the first time the hyperparameter optimization value of the random forest algorithm with the optimized handle simulation mechanism is used to perform text classification. For the second time, the traditional random forest algorithm hyperparameters use default values for text classification.

Using a random handle simulation algorithm to perform hyperparameter handle simulation, the experimental results of random handle simulation are shown in Figure 9. Figure 9 is a scatter plot of three-dimensional coordinates, where the  $x$  and  $y$  coordinates represent the number of decision trees and the number of features in the feature attribute subset, respectively, and the  $z$  coordinate is the evaluation index of the experimental algorithm, that is, the average score of the scores on the test set. Each point corresponds to the score on the test set under the condition that the hyperparameter value is taken.

It can be seen from Figure 9 that, in the case where the number of decision trees and the number of features in the feature attribute subset are different value combinations, the score on the test set is significantly different. When the number of decision trees is small and the number of features in the subset of feature attributes is large, the score on the test set is significantly lower. If only one of the number of decision trees and the number of features in the subset of feature attributes is considered, basically the greater the number of decision trees, the higher the score on the test set; the smaller the number of features in the subset of feature attributes, the higher the score on the test set high. According to the results, the best five sets of experimental hyperparameter values are shown in Figure 10.

Results of random handle simulation are analyzed to determine the number of grid handle simulation. By analyzing the five sets of hyperparameter value combinations in Figure 10, it can be seen that the first set of experimental results is significantly better than the other four sets. According to the algorithm proposed above, the hyperparameter values of the five sets of results are similar in performance. The gap is large, and a grid handle simulation is performed around the optimal value. That is, only one grid handle simulation is needed near  $k = 279$  and  $m = 3$ . Carry on the grid handle simulation to get the final hyperparameter value. Set the hyperparameter range of the grid handles simulation algorithm. The range of  $k$  is  $250 \leq k \leq 340$ , the step size is 10, the range of  $m$  is  $2 \leq m \leq 6$ , and the step size is 1. The handle simulation result of the grid handles simulation algorithm, as shown in Figure 11. Figure 11 is a graph under three-dimensional coordinates, where the  $x$  coordinate and the  $y$  coordinate represent the number of decision trees and the number of features in the feature attribute subset, and the  $z$  coordinate is the evaluation index of the experimental algorithm, that is, the average score of the score on the test set. The points on each grid represent the

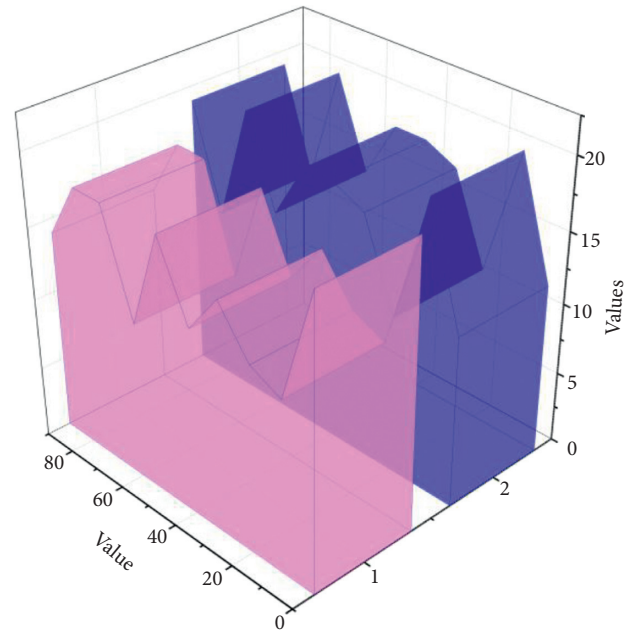


FIGURE 9: Random handle simulation results.

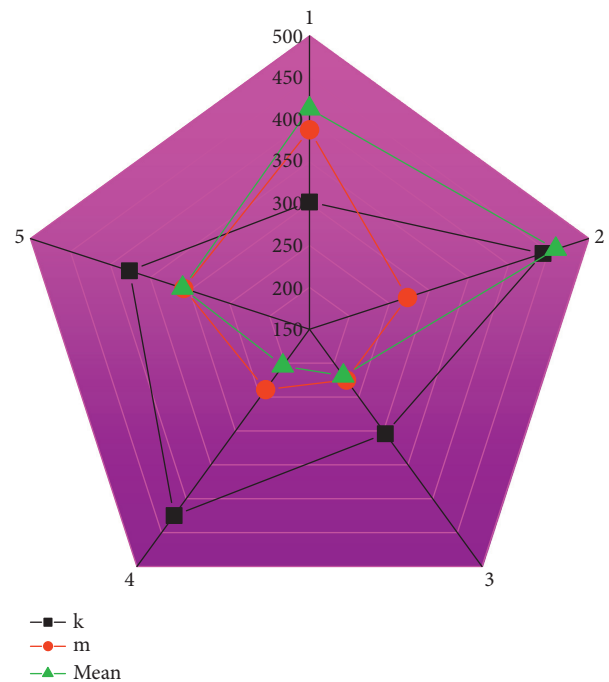


FIGURE 10: The best five sets of experimental parameter values and experimental results.

score of the algorithm on the test set when the number of decision trees and the number of features in the subset of feature attributes are the corresponding horizontal and vertical coordinates. Connect each point in a line to form a grid-like graph. Different colors are used to distinguish the score of different test sets. The transition from dark blue to red corresponds to the high to low score on the test set. According to the results of the grid handle simulation, the

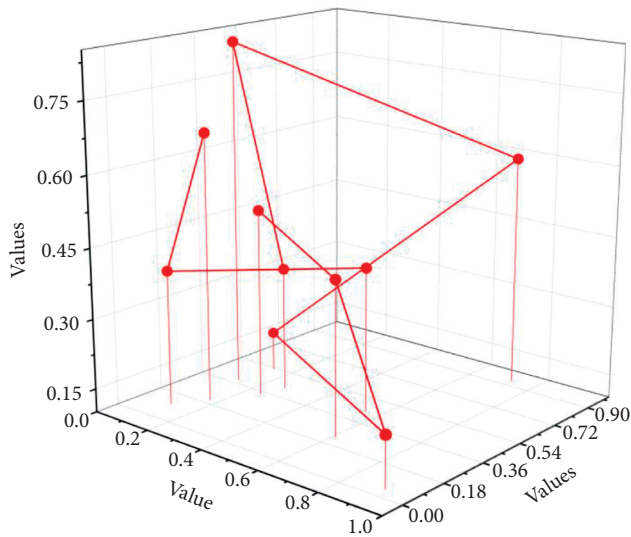


FIGURE 11: Grid handle simulation results.

optimal value of the hyperparameters of the algorithm is obtained,  $k = 270$ ,  $m = 2$ , and the mean value of the score is the highest at 0.8362.

After text-processing the data set and expressing it as a data set suitable for training the classifier, the random forest algorithm that performs the traditional random forest algorithm and the optimization method of the handle simulation mechanism based on weighted handle simulation proposed in the paper are used for text classification, experimental algorithm comparison experiment. Using time and algorithm prediction accuracy as indicators to evaluate the performance of the two algorithms, the experimental results can be concluded by analyzing the experimental results. The random forest algorithm proposed in this paper based on the optimization method of the handle simulation mechanism based on weighted handle simulation has more advantages than the traditional random forest classification algorithm.

#### 4. Conclusion

In this paper, through the improved decision tree algorithm, multimedia English-assisted handle parsing and simulation are studied. After analyzing the research status, the vacancies in the current price segment pronunciation error detection research field are summarized, and a classification error detection model method based on machine learning algorithms is proposed. To solve the pronunciation problem caused by the learner's pronunciation is not standard, so as to make more intuitive suggestions for the learner's pronunciation. A random forest model that optimizes the handle simulation mechanism is proposed. Handle simulation weight of each decision tree is obtained by the classification correct rate of the decision tree. The probability that the output sample belongs to each class when deciding on the sample result is determined. The output of the decision tree is used. The class probabilities and the weight of the decision tree are weighted to simulate the handles to

obtain the handle simulation of samples belonging to each class, and the final classification result is obtained. Design text classification experiments to compare the traditional random forest algorithm and the random forest algorithm of the optimized handle simulation mechanism proposed in this paper. The analysis of the experimental results proves that the random forest algorithm of the optimized handle simulation mechanism improves when the acceptable time complexity increases. The handle simulation ability of the algorithm is improved, and the purpose of improving the classification performance is achieved.

#### Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

#### Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### References

- [1] Y. Huo, "Analysis of intelligent evaluation algorithm based on English diagnostic system," *Cluster Computing*, vol. 22, no. 6, pp. 13821–13826, 2019.
- [2] Y. Choi and H. Lee, "Data properties and the performance of sentiment classification for electronic commerce applications," *Information Systems Frontiers*, vol. 19, no. 5, pp. 993–1012, 2017.
- [3] A. Kumar and A. Jaiswal, "Swarm intelligence based optimal feature selection for enhanced predictive sentiment accuracy on twitter," *Multimedia Tools and Applications*, vol. 78, no. 20, pp. 29529–29553, 2019.
- [4] B. Münzer, K. Schoeffmann, and L. Böszörményi, "Content-based processing and analysis of endoscopic images and videos: a survey," *Multimedia Tools and Applications*, vol. 77, no. 1, pp. 1323–1362, 2018.
- [5] Z. Li and Y. Zhang, "Hierarchical evaluation algorithm of logistics carrying capacity based on transfer learning in multimedia environment," *Multimedia Tools and Applications*, vol. 78, no. 4, pp. 4481–4501, 2019.
- [6] H. Ji, Y. Yun, S. Lee, K. Kim, and H. Lim, "An adaptable UI/UX considering user's cognitive and behavior information in distributed environment," *Cluster Computing*, vol. 21, no. 1, pp. 1045–1058, 2018.
- [7] L. Pan, "A big data-based data mining tool for physical education and technical and tactical analysis," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 14, no. 22, pp. 220–231, 2019.
- [8] B. Cai, L. Huang, and M. Xie, "Bayesian networks in fault diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 5, pp. 2227–2240, 2017.
- [9] L. Deng, D. Li, X. Yao, D. Cox, and H. Wang, "Mobile network intrusion detection for IoT system based on transfer learning algorithm," *Cluster Computing*, vol. 22, no. 4, pp. 9889–9904, 2019.
- [10] Z. Xu, "Dynamic monitoring and management system for land resource based on parallel network algorithm and remote sensing," *Journal of Intelligent & Fuzzy Systems*, vol. 37, no. 1, pp. 249–262, 2019.

- [11] Z. Lai and Q. Chen, "Reconstructing granular particles from X-ray computed tomography using the TWS machine learning tool and the level set method," *Acta Geotechnica*, vol. 14, no. 1, pp. 1–18, 2019.
- [12] Z. Subecz, "Event detection and classification in natural texts," *GRADUS*, vol. 6, no. 1, pp. 16–21, 2019.
- [13] Z. Ma, Y. Lai, W. B. Kleijn, Y.-Z. Song, L. Wang, and J. Guo, "Variational Bayesian learning for Dirichlet process mixture of inverted Dirichlet distributions in non-Gaussian image feature modeling," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 2, pp. 449–463, 2018.
- [14] M. J. W. McLarnon and T. A. O'Neill, "Extensions of auxiliary variable approaches for the investigation of mediation, moderation, and conditional effects in mixture models," *Organizational Research Methods*, vol. 21, no. 4, pp. 955–982, 2018.
- [15] M. A. Hariri-Ardebili and F. Pourkamali-Anaraki, "Support vector machine based reliability analysis of concrete dams," *Soil Dynamics and Earthquake Engineering*, vol. 104, pp. 276–295, 2018.
- [16] R. Yan, "Researches on hybrid algorithm for moving target detection and tracking in sports video," *Cluster Computing*, vol. 22, no. 2, pp. 3543–3552, 2019.
- [17] J. Ma and T. W. S. Chow, "Topic-based algorithm for multilabel learning with missing labels," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 7, pp. 2138–2152, 2018.
- [18] Y. Wu and Q. Ji, "Facial landmark detection: a literature survey," *International Journal of Computer Vision*, vol. 127, no. 2, pp. 115–142, 2019.
- [19] M. B. Mustafa, M. A. M. Yusoof, Z. M. Don, and M. Malekzadeh, "Speech emotion recognition research: an analysis of research focus," *International Journal of Speech Technology*, vol. 21, no. 1, pp. 137–156, 2018.
- [20] Y. Li, L. Zhang, Y. Xu, Y. Yao, R. Y. K. Lau, and Y. Wu, "Enhancing binary classification by modeling uncertain boundary in three-way decisions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 7, pp. 1438–1451, 2017.
- [21] E. Gundogdu, H. Ozkan, and A. A. Alatan, "Extending correlation filter-based visual tracking by tree-structured ensemble and spatial windowing," *IEEE Transactions on Image Processing*, vol. 26, no. 11, pp. 5270–5283, 2017.
- [22] L. Bilke, B. Flemisch, T. Kalbacher, O. Kolditz, R. Helmig, and T. Nagel, "Development of open-source porous media simulators: principles and experiences," *Transport in Porous Media*, vol. 130, no. 1, pp. 337–361, 2019.

## Research Article

# Emotion Recognition of Students Based on Facial Expressions in Online Education Based on the Perspective of Computer Simulation

Weiqing Wang,<sup>1</sup> Kunliang Xu,<sup>1</sup> Hongli Niu,<sup>1</sup> and Xiangrong Miao <sup>2</sup>

<sup>1</sup>School of Economics and Management, University of Science and Technology Beijing, Beijing 100083, China

<sup>2</sup>Beijing Energy Conservation and Environmental Protection Center, Beijing 101199, China

Correspondence should be addressed to Xiangrong Miao; [miaoxr@bjbeec.cn](mailto:miaoxr@bjbeec.cn)

Received 14 July 2020; Revised 26 August 2020; Accepted 28 August 2020; Published 11 September 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Weiqing Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Online education has developed rapidly due to its irreplaceable convenience. Under the severe circumstances caused by COVID-19 recently, many schools around the world have delayed opening and adopted online education as one of the main teaching methods. However, the efficiency of online classes has long been questioned. Compared with traditional face-to-face classes, there is a lack of direct, timely, and effective communication and feedback between teachers and students in the online courses. Previous studies have shown that there is a close and stable relationship between a person's facial expressions and emotions generally. From the perspective of computer simulation, a framework combining a face expression recognition (FER) algorithm with online courses platforms is proposed in this work. The cameras in the devices are used to collect students' face images, and the facial expressions are analyzed and classified into 8 kinds of emotions by the FER algorithm. An online course containing 27 students conducted on Tencent Meeting is used to test the proposed method, and the result proved that this method performs robustly in different environments. This framework can also be applied to other similar scenarios such as online meetings.

## 1. Introduction

Facial expression is one of the most powerful, natural, and universal signals for human beings to convey their emotional states and intentions regardless of national borders, race, and gender [1, 2], and there were multitudinous related applications such as the health management [3], aided driving [4, 5], and others [6–9]. In the earlier researches on facial expressions of emotion, Ekman and Friesen argued that human beings perceive certain basic emotions in the same way regardless of their cultural background, and they defined the typical facial expressions into 6 categories: anger, disgust, fear, happiness, sadness, and surprise [10, 11]. Based on studies of Ekman and Friesen, Ekman and Heider [12–14], Matsumoto [15] provided enough proofs for another universal facial expression, contempt. Additionally, FER2013 [16], a large-scale and unconstrained database introduced in the ICML 2013 Challenges in Representation Learning,

labeled its facial images into anger, disgust, fear, happiness, sadness, surprise, and neutral, which has been widely used in designing facial expression recognition (FER) systems. In subsequent researches, although researchers introduced many models that can provide a wider range of emotions to deal with the complexity and subtlety of facial expressions [17–20], the classification that describes basic emotions discretely is still the most widely used method in FER due to its generality and intuitive definition of facial expressions [21], and Figure 1 displays the 8 basic facial expression phenotypes from datasets CK+ [22] and FER2013 [16].

For determining facial expressions, Ekman and Friesen [23] proposed a Facial Action Coding System (FACS) for determining facial expressions, which is based on a fact that expressions result from the change of facial parts. With the assistance of the computers, more advanced methods have been proposed during the last decades [24], and the feature points can be seen in Figure 2.



FIGURE 1: Basic facial expression phenotypes from CK+ and FER2013: from left to right are anger, contempt, disgust, fear, happiness, neutral, sadness, and surprise.

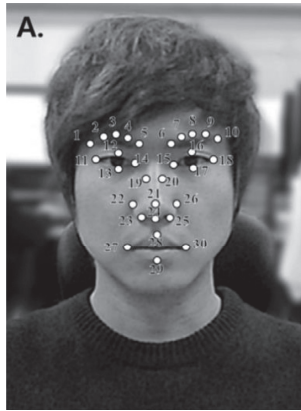


FIGURE 2: Facial analysis points used for facial expression [25].

With the development of artificial intelligence and deep learning, numerous FER algorithms have been proposed to deal with the expression information in facial representations, which has improved the accuracy of recognition gradually and achieved better performance than traditional methods [26, 27]. The tasks of FER can be mainly divided into two categories: static images (represented by photographs) [28–30] and dynamic sequence (represented by videos) [31–33] that take the dynamic relationship between the continuously changing images into account and therefore pose additional challenges than the former. In addition to the vision-based methods, other biometric techniques [34, 35] can also be adopted to assist the recognition of expression.

Sufficient labeled training databases that include as many variations of the populations and environments as possible are important for researchers to design and test a FER model or system; the existing databases are mainly divided into controlled and uncontrolled. On one hand, the controlled databases, represented by CK+ [22], Jaffe [36], and MMI [37], are collected from laboratory environments with sufficient light and simple backgrounds. Nowadays, because most actual scenes are complex and changeable due to factors such as lighting, FER in laboratory or controlled environments is generally considered to be of little practical significance and used mainly for the proof of concept of features extraction and classification methods. On the other hand, the uncontrolled databases, such as FER2013 [15] and AFEW [38], are collected from complex environments with vastly different backgrounds, occlusions, and illuminations; these scenes are more similar to the actual situations and have been increasingly used in more and more researches.

Limited by the hardware and insufficient processing capability, the majority of the traditional methods for FER

employed hand-craft features or shallow learning, such as local binary patterns (LBP) [28] and nonnegative matrix factorization (NMF) [39]. With the development of processing capabilities and computer simulation, all kinds of machine learning algorithms, such as Artificial Neural Networks (ANNs), Support Vector Machines (SVM), and Bayesian classifiers, were applied to FER, and the high accuracy has been verified in controlled environments so that the faces can be detected effectively. However, these methods were weak in generalization ability while this is the key to evaluate the practicality of a model [40]. Deep learning algorithms can solve this problem, and it is also robust in the uncontrolled environments. Recent works have shown that convolutional neural networks (CNNs), because of their effectiveness in feature extraction and classification tasks, performed well in addressing the computer vision problems especially in FER [41, 42], and numerous models based on CNN structure are proposed constantly and have achieved better results than previous methods. Simonyan and Zisserman [43] adopted an architecture of very small ( $3 \times 3$ ) convolution filters to conduct a comprehensive evaluation of networks with increasing depth and the two best-performing ConvNet models were available publicly to facilitate the further research in this field. By increasing the depth and width of the network while keeping the computational budget constant, Szegedy et al. [44] introduced a deep convolutional neural network architecture named “Inception” in which the utilization of the computing resources can be improved significantly, and Jahandad et al. [45] worked on 2 convolutional neural network architectures (Inception-v1 and Inception-v3) based on “Inception” and proved that these 2 models performed better than others, and Inception-v1 with 22-layer-deep network performed better than 42-layer-deep Inception-v3 network when facing low-resolution input images and 2D images of signatures; however, Inception-v3 outperformed in ImageNet challenge. The general trend of neural networks is to increase the depth of the network and the width of layer. In theory, the deeper the neural network models, the stronger the learning capabilities, but the more difficult to train. He et al. [46] proposed a residual learning framework to reduce training difficulty of deeper networks and proved thoroughly that these residual networks are easier to optimize while increasing accuracy from the considerably increased depth. In addition, a part of researchers proposed that the accuracy of recognition can be further improved by combining CNNs with recurrent neural networks (RNNs) in which the CNNs are adopted as the inputs to RNNs [47, 48].

During the past decades, online education has developed rapidly whether at universities or training institutions [49], which offers potential application opportunities for FER.



Significantly different from the traditional face-to-face courses, online courses are often considered of less constraining force and effective communication, which will inevitably lead to faculty's suspicions towards this novel educational method [50, 51] while there are several studies that argue that the students' learning outcomes achieved by online education may be comparable to traditional face-to-face courses [52, 53], except for the skills that require optimum precision and a greater degree of haptic awareness [54]. It is undeniable that the rapid growth of online education can effectively provide the convenience and flexibility for more students, so it also has broad development space in the future; therefore, how to ensure that students keep the same level of concentration and learning efficiency as the traditional courses during online education is critical to promote the further development of online education.

In brief, the main contribution of this paper is as follows. By combining the existing online education platforms with facial expression recognition model based on the architecture of convolutional neural network, this work proposed a framework that enables real-time monitoring of students' emotions in online courses and ensures that the feedback expressed by facial expression can be provided to teachers timely, so that they can flexibly adjust the teaching programs and ultimately improve the quality and efficiency of online education.

## 2. Proposed Framework

The framework mainly consists of two parts: online courses platforms, in this paper we took Tencent Meeting as an example for mode testing, and a deep learning model based on CNN, inspired by Kuo et al. [27], before which it is noted that the original images collected from online courses need to be preprocessed, including face detection, alignment, rotation, and resize, according to the different elements in the original images. Figure 3 exhibits the process of the FER, and the detailed steps of the proposed framework are as follows: first, the cameras built in the electronic devices are utilized to capture the facial images of the attending students. Second, the facial expression recognition algorithm trained by the standard facial expression database is employed to detect the faces and classify the facial expressions in terms of anger, disgust, fear, happiness, sadness, surprise, contempt, and neutral. Third, the histogram of probability distribution about the expression is plotted and provided for the teacher so that the teaching plan can be adjusted timely.

*2.1. Online Education Platforms.* The advances in technological delivery modalities have spawned a large number of online education platforms and greatly improved the flexibility of education, enabling teachers to adopt diverse technical methods to assist teaching without worrying about the limit on the number of students in traditional classroom-based course, and students in different regions can communicate in real time without having to consider the traffic and other issues. The same teaching materials as traditional

classes can be uploaded to these platforms for students' reference. Currently, in platforms that have online teaching functions, such as DingTalk, Zoom, and Rain classroom, teachers can adopt the method of video meeting and take advantage of the camera built in devices to capture and recognize students' facial expressions in real time. The captured images will be preprocessed and then used as the input of CNN.

*2.2. The Preprocessing Based on IntraFace.* Effective preprocessing can reduce the interference of face-like objects in the background when detecting faces in an image and then standardize the face images according to the heuristic knowledge, which will effectively improve the efficiency of the deep learning model. We employed IntraFace [55], a publicly available software package that integrates algorithms for facial feature tracking, head pose estimation, facial attribute detection, etc., as the tool of preprocessing. As shown in Figure 4, IntraFace can also be used to detect multiple faces at the same time. The key features of each face including eyebrows, eyes, nose tip, and mouth can be recognized effectively, and the facial expression can be detected by rectangular outlines accordingly; these outlines are constructed by the feature points at the edge of every face, including the uppermost and the lowermost, which determined the vertical width, and the rightmost and the leftmost, which determined the horizontal width of the face image. In order to prevent the omission of facial information while reducing the noise of background, we enlarge the rectangular outlines by 1.05 times to cover more facial content. Furthermore, considering that the size of images input into the learning model is preset to  $48 \times 48$ , the detected images will be rotated with nose tip as the center and resized appropriately to make it consistent with the input size.

*2.3. The Learning Model Based on CNN.* The architecture of the applied deep learning model based on CNN is illustrated in Figure 5, which referred to the research results proposed by Kuo et al. [27], and the prior performance of this model in FER over the other similar has also been proven. After a convolutional layer of 32 feature maps, the input layer is followed by 2 blocks, which consists of 2 convolutional layers and 1 max-pooling layer with 64 feature maps separately. And the size of kernels in the first convolutional layer is set to  $3 \times 3$ , the second is  $5 \times 5$ , the max-pooling layers both consist of a kernel of size  $2 \times 2$  and stride 2, and as a consequence, the input image will be compressed to a quarter. And there are 2 following fully connected layers of 2048 and 1024 neurons, respectively, in which Rectified Linear Units (ReLUs) [56–59] are adopted as the activation function. In order to prevent overfitting, a Dropout is added after each of the 2 fully connected layers, which will release a part of neurons according to the presetting drop-probability; in this paper, the 2 values are both set to 0.5. The following output layer is composed of 8 units, and softmax [60] is adopted as the activation function to classify the expressions examined in terms of anger, disgust, fear, happiness, sadness, surprise, contempt, and neutral.

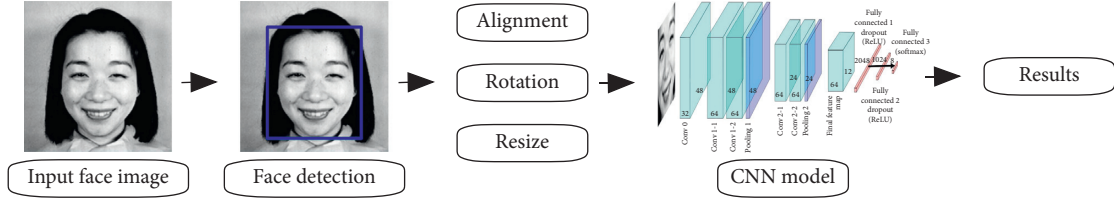


FIGURE 3: The process of FER.

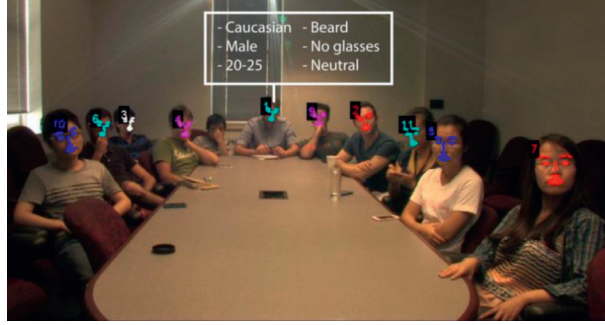


FIGURE 4: Automatic output of IntraFace to measure audience reaction [55].

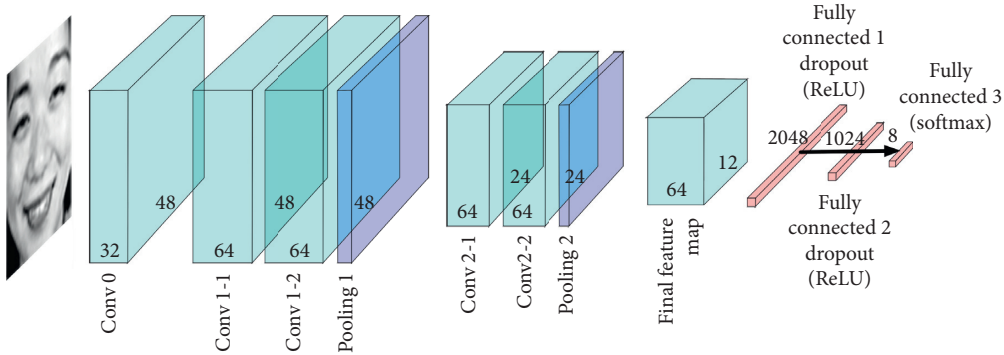


FIGURE 5: The structure of the applied CNN model.

The proposed model was trained by databases Jaffe, CK+, and FER2013 that include the above 8 basic expressions at the same time. Because the small FER databases usually only contain a few hundreds of images, obviously this is not enough for model training; we adopt online augmentation strategy with both horizontal flipping and random shifting to increase the images of training sets. More details about the CNN model are given in Table 1.

And in this model, the output size  $N$  of each convolutional layer can be formulated as

$$N = \frac{I - F + 2P}{S} + 1, \quad (1)$$

where  $I, F, P$ , and  $S$  denote the input size, kernel size, padding size, and stride size, respectively.

In each max-pooling layer, the padding size is 0, and the output size can also be expressed as

$$N = \frac{I - F}{S} + 1. \quad (2)$$

Rectified Linear Units (ReLU) are adopted as the activation function in the convolutional layers and max-pooling layers to avoid gradient explosion and ensure faster convergence speed during the back-propagation operation, which can be formulated as

$$f(x) = \begin{cases} 0, & x \leq 0; \\ x, & x > 0. \end{cases} \quad (3)$$

Softmax is used as the activation function in the output layer, the input of which is the matrix  $z_i = w_i^T x + b_i$

TABLE 1: The parameters of the applied CNN model.

| Layers            | Kernel number | Kernel size   | Steps | Image size               |
|-------------------|---------------|---------------|-------|--------------------------|
| Input             | 0             | 0             | None  | $48 \times 48 \times 1$  |
| Convolutional 0   | 32            | $1 \times 1$  | 1     | $48 \times 48 \times 32$ |
| Convolutional 1-1 | 64            | $3 \times 3$  | 1     | $48 \times 48 \times 64$ |
| Convolutional 1-2 | 64            | $5 \times 5$  | 1     | $48 \times 48 \times 64$ |
| Pooling 1         | 0             | $2 \times 2$  | 2     | $24 \times 24 \times 64$ |
| Convolutional 2-1 | 64            | $3 \times 3$  | 1     | $24 \times 24 \times 64$ |
| Convolutional 2-2 | 64            | $5 \times 5$  | 1     | $24 \times 24 \times 64$ |
| Pooling 2         | 0             | $2 \times 2$  | 2     | $12 \times 12 \times 64$ |
| Fully connected 1 |               | Dropout = 0.5 |       | $1 \times 1 \times 2048$ |
| Fully connected 2 |               | Dropout = 0.5 |       | $1 \times 1 \times 1024$ |
| Output            | 0             | 0             | None  | $1 \times 1 \times 8$    |

output from the fully connected layers. The formulation is given as

$$S_i = \frac{e^{z_i}}{\sum_{k=1}^K e^{z_k}}, \quad (4)$$

where  $K$  represents the output dimensions of the layer, meaning there are  $K$  kinds of results, and  $S_i$  represents the probability of result  $i$ ,  $i = 1, 2, \dots, K$ .

And Softmax loss, which is used for gradient derivation and update, can be calculated by the following:

$$L = - \sum_{i=1}^K y_i \log S_i,$$

$$\frac{\partial L}{\partial z_k} = \frac{\partial L}{\partial S_i} \frac{\partial S_i}{\partial z_k},$$

$$\frac{\partial L}{\partial S_i} = \sum_{i=1}^K \left( -\frac{y_i}{S_i} \right),$$

$$\frac{\partial S_i}{\partial z_k} = \begin{cases} \frac{e^{z_i} \sum_{k=1}^K e^{z_k} - e^{z_i} e^{z_k}}{(\sum_{k=1}^K e^{z_k})^2} = S_i(1 - S_k), & i = k; \\ \frac{e^{z_i} \cdot e^{z_k}}{(\sum_{k=1}^K e^{z_k})^2} = -S_i S_k, & i \neq k, \end{cases} \quad (5)$$

$$\begin{aligned} \frac{\partial L}{\partial z_k} &= \frac{\partial L}{\partial S_i} \frac{\partial S_i}{\partial z_k} = - \sum_{i=1}^K \frac{y_i}{S_i} \cdot \frac{\partial S_i}{\partial z_k} \\ &= -S_k(1 - S_k) \cdot \frac{y_k}{S_k} + \sum_{i \neq k} \left( -\frac{y_i}{S_i} \right) \cdot (-S_i S_k) \\ &= -y_k + y_k S_k + \sum_{i \neq k} y_i S_k = S_k - y_k, \end{aligned}$$

where  $L$  denotes the loss function and  $y_i$  is the label variable, the value of which is 1 or 0 according to whether the output is consistent with the actual value.

### 3. Experiment and Results

In order to test the performance of the proposed framework in practical applications, we captured an image that includes 27 people from an online meeting held on Tencent Meeting and input it into the CNN model. This image is taken before the end of the meeting; the moderator was making a concluding speech in a pleasant atmosphere. In addition, everyone was told that the meeting was coming to an end, according to the experiment conducted by Tonguç and Ozkara [61], students' happiness will be significantly improved within a few minutes before the end of a lecture, so under the similar circumstance, it can be inferred that the emotions presented by most of the faces in this image are happy or neutral. Figure 6 shows the input (left) and output (right) images of the CNN model. It can be seen clearly from the result that all the faces were recognized and marked by the rectangular outlines, and the responding facial expressions were also labeled. In the total 27 faces, 10 were labeled "happy," 15 were labeled "neutral," and 2 were labeled "sad," noting that the 2<sup>nd</sup> image in the last line and the 3<sup>rd</sup> image in the 4<sup>th</sup> line from last, marked by red outlines, were not detected by the outlines precisely; the reason may be that the 2 face images are so incomplete that the features presented are too insufficient to recognize. Figure 7 shows the probability distribution histogram of emotions, from which we can observe the overall emotions intuitively and judge the emotional state of class accordingly. It is worth noting that the probability of happiness is significantly higher than that of neutral in this figure, while the faces labeled "happy" are less than "neutral" as exhibited in Figure 6. The difference can be explained as follows: there may be features of multiple expressions on a face at the same time; the expression presented on this face will be labeled according to the most likely expression decided by these features, but the overall expression of an image including multiple faces is decided by the sum of various expression features contained in each face. In some faces that are marked as "happy," the probability of happiness may be much higher than neutral, while in some faces that are marked as "neutral," the probability of happiness may be only slightly lower than neutral. Overall, the result of this experiment can provide favorable support for the performance of the model when applied to real environment.

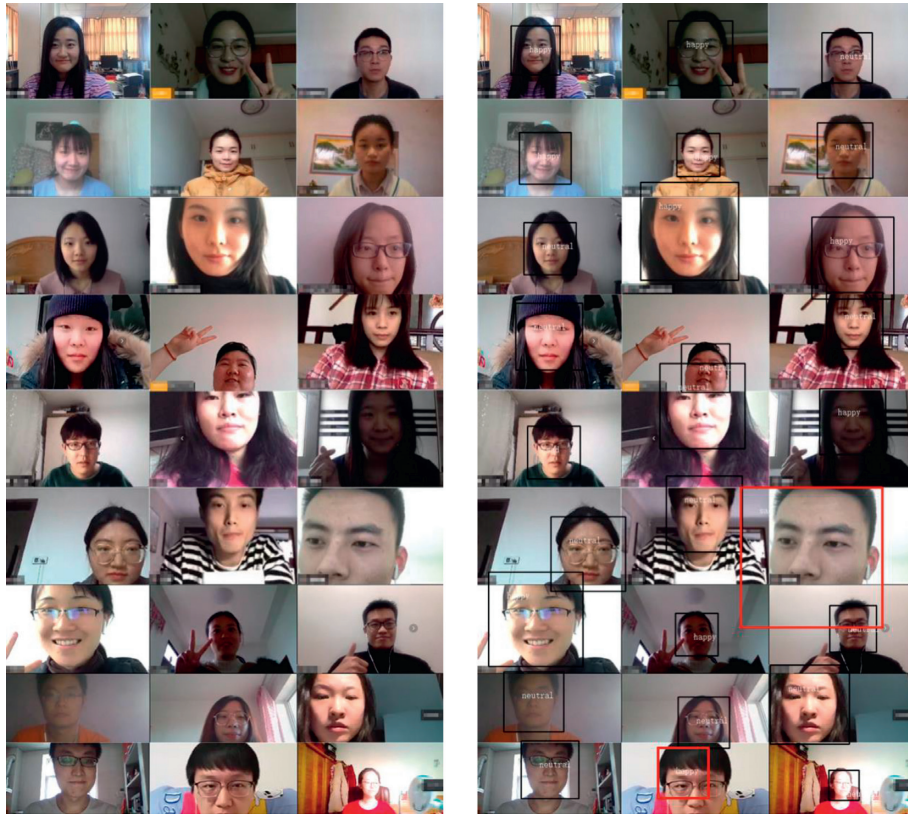


FIGURE 6: The input (a) and output (b) images of the FER model.

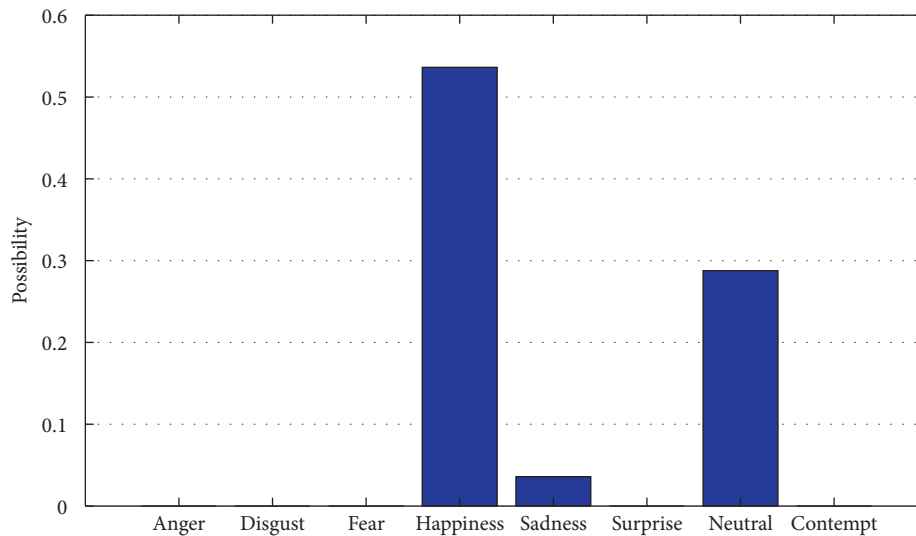


FIGURE 7: The probability distribution of emotions.

#### 4. Conclusion and Discussion

In this study, by combining the online courses platforms and a compact deep learning model based on the architecture of CNN, we construct a framework to analyze students' emotions according to their facial expressions from the perspective of computer simulation. The overall result can be

presented in a histogram intuitively, and teachers can adjust their teaching strategies accordingly to improve the efficiency of online teaching.

With reference to the studies of Ekman et al. and FER2013, the emotions were classified into anger, disgust, fear, happiness, sadness, surprise, contempt, and neutral in the proposed framework. To verify the applicability of this

framework in a real environment, we captured an image including the facial images of all participants at one time in a real online meeting; there were 12 participants in this meeting, and the captured time was determined at the end of meeting. A total of 12 faces were captured, of which 11 were effectively recognizable faces that contain enough feature points. By inputting this image into the applied CNN model, we obtained the emotional tags for each valid face and got the overall emotion at that time. It has been proved that the framework has good applicability in practical activities and plays a positive role in solving the problems, such as the lack of binding force on students, and teachers cannot get timely feedback. Ultimately, it will contribute to improving the quality of online education.

Despite the above benefits, there is still much room for improvement in this framework and its applications. From the perspective of technologies, with the development of computer simulation, algorithms with better performance and shorter operation time, including preprocessing and deep learning models, will be continuously developed over time. For instance, the image preprocessing contains face detection, alignment, rotation, and resize, but when facing problems, such as backlight, shadows, and facial incompleteness, caused by complex environments, these current methods are always powerless, and these shortcomings may be solved in the future. What is more, although the CNN model in the proposed framework currently performs well, it will be replaced by models with higher learning capabilities and higher classification accuracy in the future. In order to ensure the competitiveness of the framework in a longer period, it needs to be adjusted and maintained regularly, and more advanced algorithms and technologies should be adopted to update it.

In addition, with a large number of participants in the online courses, we have no way to ensure that everyone keeps the high level of concentration, and then students' expressions may not fully represent their emotions due to these subjective factors. Taking measures like setting thresholds can filter out some invalid information and highlight the main emotions in the image. Finally, the teaching efficiency can be improved as a result.

### Data Availability

The data used in this manuscript can be accessed by readers via the authors' BaiduPan at <https://pan.baidu.com/s/1dbKUfeKp5joeYh4wSOU7Qw> with the extraction code "qin6."

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Acknowledgments

This work was supported by the Study on Influence of Chinese Stock Market under Economic Uncertainty (No. FRF-DF-20-11).

### References

- [1] C. Darwin and P. Prodger, *The Expression of the Emotions in Man and Animals*, Oxford University Press, Oxford, MA, USA, 1998.
- [2] Y.-I. Tian, T. Kanade, and J. F. Cohn, "Recognizing action units for facial expression analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 97–115, 2001.
- [3] R. Donoso, C. San Martín, and G. Hermosilla, "Reduced iso-thermal feature set for long wave infrared (LWIR) face recognition," *Infrared Physics & Technology*, vol. 83, pp. 114–123, 2017.
- [4] T. Liu, H. Liu, Z. Chen et al., "FBRDLR: fast blind reconstruction approach with dictionary learning regularization for infrared microscopy spectra," *Infrared Physics & Technology*, vol. 90, pp. 101–109, 2018.
- [5] Z. Huang, H. Fang, Q. Li et al., "Optical remote sensing image enhancement with weak structure preservation via spatially adaptive gamma correction," *Infrared Physics & Technology*, vol. 94, pp. 38–47, 2018.
- [6] Y. Bi, M. Lv, Y. Wei, N. Guan, and W. Yi, "Multi-feature fusion for thermal face recognition," *Infrared Physics & Technology*, vol. 77, pp. 366–374, 2016.
- [7] H. Liu, Z. Zhang, S. Liu, J. Shu, T. Liu, and T. Zhang, "Blind spectrum reconstruction algorithm with  $L_0$ -sparse representation," *Measurement Science and Technology*, vol. 26, no. 8, pp. 085501–085507, 2015.
- [8] H. Wu, Y. Liu, L. Qiu, and Y. Liu, "Online judge system and its applications in C language teaching," in *Proceedings of the International Symposium on Educational Technology (ISET)*, pp. 57–60, Beijing, China, July 2016.
- [9] T. Liu, Z. Chen, H. Liu, Z. Zhang, and Y. Chen, "Multi-modal hand gesture designing in multi-screen touchable teaching system for human-computer interaction," in *Proceedings of the Second International Conference on Advances in Image Processing*, pp. 100–109, Chengdu China, June 2018.
- [10] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *Journal of Personality and Social Psychology*, vol. 17, no. 2, pp. 124–129, 1971.
- [11] P. Ekman, "Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique," *Psychological Bulletin*, vol. 115, no. 2, pp. 268–287, 1994.
- [12] P. Ekman and W. V. Friesen, "A new pan-cultural facial expression of emotion," *Motivation and Emotion*, vol. 10, no. 2, pp. 159–168, 1986.
- [13] P. Ekman and W. V. Friesen, "Who knows what about contempt: a reply to Izard and Haynes," *Motivation and Emotion*, vol. 12, no. 1, pp. 17–22, 1988.
- [14] P. Ekman and K. G. Heider, "The universality of a contempt expression: a replication," *Motivation and Emotion*, vol. 12, no. 3, pp. 303–308, 1988.
- [15] D. Matsumoto, "More evidence for the universality of a contempt expression," *Motivation and Emotion*, vol. 16, no. 4, pp. 363–368, 1992.
- [16] I. J. Goodfellow, D. Erhan, P. L. Carrier et al., "Challenges in representation learning: a report on three machine learning contests," *Neural Information Processing*, Springer, Berlin, Germany, pp. 117–124, 2013.
- [17] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: audio, visual, and spontaneous expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 39–58, 2009.
- [18] E. Sariyanidi, H. Gunes, and A. Cavallaro, "Automatic analysis of facial affect: a survey of registration, representation, and

- recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 6, pp. 1113–1133, 2015.
- [19] B. Martinez and M. F. Valstar, "Advances, challenges, and opportunities in automatic facial expression recognition," in *Advances in Face Detection and Facial Image Analysis*, pp. 63–100, Springer, Cham, Switzerland, 2016.
- [20] H. Gunes and B. Schuller, "Categorical and dimensional affect analysis in continuous input: current trends and future directions," *Image and Vision Computing*, vol. 31, no. 2, pp. 120–136, 2013.
- [21] S. Li and W. Deng, "Deep facial expression recognition: a survey," *IEEE Transactions on Affective Computing*, In press.
- [22] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression," in *Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, pp. 94–101, San Francisco, CA, USA, July 2010.
- [23] P. Ekman and W. V. Friesen, "Measuring facial movement," *Environmental Psychology and Nonverbal Behavior*, vol. 1, no. 1, pp. 56–75, 1976.
- [24] S. Bayrakdar, D. Akgün, and İ. Yücedağ, "A survey on automatic analysis of facial expressions," *Sakarya University Journal of Science*, vol. 20, no. 2, pp. 383–398, 2016.
- [25] Y. Kim, S. Kang, S. Lee et al., "Efficiently detecting outlying behavior in video-game players," *PeerJ*, vol. 3, p. e1502, 2015.
- [26] A. Agrawal and N. Mittal, "Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy," *The Visual Computer*, vol. 36, no. 2, pp. 405–412, 2020.
- [27] C. M. Kuo, S. H. Lai, and M. Sarkis, "A compact deep learning model for robust facial expression recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 2202–2208, Salt Lake City, UT, USA, June 2018.
- [28] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: a comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.
- [29] P. Liu, S. Han, Z. Meng, and Y. Tong, "Facial expression recognition via a boosted deep belief network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1805–1812, Columbus, OH, USA, June 2014.
- [30] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *Proceedings of the IEEE Winter Conference Applications of Computer Vision (WACV)*, pp. 1–10, Lake Placid, NY, USA, March 2016.
- [31] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 915–928, 2007.
- [32] H. Jung, S. Lee, J. Yim, S. Park, and J. Kim, "Joint fine-tuning in deep neural networks for facial expression recognition," in *Proceedings of the IEEE International Conference Computer Vision (ICCV)*, pp. 2983–2991, Santiago, Chile, December 2015.
- [33] X. Zhao, X. Liang, L. Liu et al., "Peak-piloted deep network for facial expression recognition," *Computer Vision—ECCV 2016*, Springer, Cham, Switzerland, pp. 425–442, 2016.
- [34] I. Bhardwaj, N. D. Londhe, and S. K. Kopparapu, "A spoof resistant multibiometric system based on the physiological and behavioral characteristics of fingerprint," *Pattern Recognition*, vol. 62, pp. 214–224, 2017.
- [35] J. Zhao, R. Ma, and X. Zhang, "Speech emotion recognition based on decision tree and improved SVM mixed model," *Transactions of Beijing Institute of Technology*, vol. 37, no. 4, pp. 386–395, 2017.
- [36] M. J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, and J. Budynek, "The Japanese female facial expression (Jaffe) database," in *Proceedings of Third International Conference on Automatic Face and Gesture Recognition*, pp. 14–16, Nara, Japan, April 1998.
- [37] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *Proceedings of the IEEE International Conference on Multimedia and Expo*, Amsterdam, Netherland, July 2005.
- [38] A. Dhall, R. Goecke, S. Lucey, and T. Gedeon, "Collecting large, richly annotated facial-expression databases from movies," *IEEE Multimedia*, vol. 19, no. 3, pp. 34–41, 2012.
- [39] R. Zhi, M. Flierl, Q. Ruan, and W. B. Kleijn, "Graph-preserving sparse nonnegative matrix factorization with application to facial expression recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 41, no. 1, pp. 38–52, 2011.
- [40] A. Dhall, R. Goecke, J. Joshi, K. Sikka, and T. Gedeon, "Emotion recognition in the wild challenge 2014: baseline, data and protocol," in *Proceedings of the 16th International Conference on Multimodal Interaction*, pp. 461–466, ACM, Istanbul Turkey, November 2014.
- [41] J. Li, K. Jin, D. Zhou, N. Kubota, and Z. Ju, "Attention mechanism-based CNN for facial expression recognition," *Neurocomputing*, vol. 411, pp. 340–350, 2020.
- [42] J. Shao and Y. Qian, "Three convolutional neural network models for facial expression recognition in the wild," *Neurocomputing*, vol. 355, pp. 82–92, 2019.
- [43] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, <https://arxiv.org/abs/1409.1556>.
- [44] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, Boston, MA, USA, June 2015.
- [45] A. Jahandad, S. M. Sam, K. Kamardin, N. N. Amir Sjarif, and N. Mohamed, "Offline signature verification using deep learning convolutional neural network (CNN) architectures GoogLeNet inception-v1 and inception-v3," *Procedia Computer Science*, vol. 161, pp. 475–483, 2019.
- [46] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV, USA, June 2016.
- [47] I. Banerjee, Y. Ling, M. C. Chen et al., "Comparative effectiveness of convolutional neural network (CNN) and recurrent neural network (RNN) architectures for radiology text report classification," *Artificial Intelligence in Medicine*, vol. 97, pp. 79–88, 2019.
- [48] H. Hasan, B. Huang, and G. Tian, "Facial expression recognition based on deep convolution long short-term memory networks of double-channel weighted mixture," *Pattern Recognition Letters*, vol. 131, pp. 128–134, 2020.
- [49] I. Allen and J. Seaman, "Digital compass learning: distance education enrollment report 2017," Babson Survey Research Group, Babson Park, MA, USA, 2017.
- [50] P. Shea, T. Bidjerano, and J. Vickers, "Faculty attitudes toward online learning: failures and successes," SUNY Research Network, Albany, NY, USA, 2016.

- [51] D. Lederman, *Conflicted Views of Technology: A Survey of Faculty Attitudes*, Inside Higher Education, Washington, WA, USA, 2018.
- [52] C. L. Cason and J. Stiller, "Performance outcomes of an online first aid and CPR course for laypersons," *Health Education Journal*, vol. 70, no. 4, pp. 458–467, 2010.
- [53] S. Maloney, M. Storr, S. Paynter, P. Morgan, and D. Ilic, "Investigating the efficacy of practical skill teaching: a pilot-study comparing three educational methods," *Advances in Health Sciences Education*, vol. 18, no. 1, pp. 71–80, 2013.
- [54] E. Dolan, E. Hancock, and A. Wareing, "An evaluation of online learning to teach practical competencies in undergraduate health science students," *The Internet and Higher Education*, vol. 24, pp. 21–25, 2015.
- [55] F. D. I. Torre, W. Chu, and X. Xiong, "IntraFace," in *Proceedings of the IEEE International Conference on Automatic Face And Gesture Recognition*, Lille, France, May 2015.
- [56] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 1097–1105, Lake Tahoe, Nevada, December 2012.
- [57] W. Wei, B. Zhou, D. Połap, and M. Woźniak, "A regional adaptive variational PDE model for computed tomography image reconstruction," *Pattern Recognition*, vol. 92, pp. 64–81, 2019.
- [58] L. Zhang, P. Shen, X. Peng et al., "Simultaneous enhancement and noise reduction of a single low-light image," *IET Image Processing*, vol. 10, no. 11, pp. 840–847, 2016.
- [59] Q. Zhu, J. Zhang, W. Wei et al., "A neuro-heuristic approach for recognition of lung diseases from X-ray images," *Expert Systems with Applications*, vol. 126, pp. 218–232, 2019.
- [60] C. S. Chin, J. Si, A. S. Clare, and M. Ma, "Intelligent image recognition system for marine fouling using Softmax transfer learning and deep convolutional neural networks," *Complexity*, vol. 2017, Article ID 5730419, 9 pages, 2017.
- [61] G. Tonguç and B. O. Ozkara, "Automatic recognition of student emotions from facial expressions during a lecture," *Computers & Education*, vol. 148, Article ID 103797, 2020.

## Research Article

# An Innovative Aerial Manipulator with Tandem Ducted Fans: Modeling, Control, and Simulation

Yibo Zhang,<sup>1,2</sup> Wei Fan ,<sup>2,3</sup> Changle Xiang,<sup>1,2</sup> Bin Xu,<sup>1,2</sup> Tianfu Ai,<sup>1</sup> Lei Yuan,<sup>4</sup> and Yang Liu<sup>4</sup>

<sup>1</sup>School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China

<sup>2</sup>Beijing Institute of Technology Chongqing Innovation Center, Chongqing 401147, China

<sup>3</sup>School of Automation, Beijing Institute of Technology, Beijing 100081, China

<sup>4</sup>Beijing Special Vehicle Research Institute, Beijing 100072, China

Correspondence should be addressed to Wei Fan; fanweixx@bit.edu.cn

Received 25 May 2020; Revised 6 August 2020; Accepted 15 August 2020; Published 8 September 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Yibo Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes an innovative ducted fan aerial manipulator, which is particularly suitable for the tasks in confined environment, where traditional multirotors and helicopters would be inaccessible. The dynamic model of the aerial manipulator is established by comprehensive mechanism and parametric frequency-domain identification. On this basis, a composite controller of the aerial platform is proposed. A basic static robust controller is designed via H-infinity synthesis to achieve basic performance, and an adaptive auxiliary loop is designed to estimate and compensate for the effect acting on the vehicle from the manipulator. The computer simulation analyses show good stability of the aerial vehicle under the manipulator motion and good tracking performance of the manipulator end effector, which verify the feasibility of the proposed aerial manipulator design and the effectiveness of the proposed controller, indicating that the system can meet the requirements of high precision operation tasks well.

## 1. Introduction

In recent years, the application of unmanned autonomous robots is increasingly diverse, and the interaction between the autonomous robot system and the environment is developed from information interaction (such as sound, light, picture) to physical interaction (replacing manpower to complete operation work) [1]. On the one hand, using autonomous robots for operation work can greatly save labor costs and significantly improve work efficiency; on the other hand, it can liberate people from heavy labor, especially the labor in dangerous and harmful environments. At present, some ground mobile robots have been applied to postearthquake rescue [2], some underwater robots have been applied to oceanic biological sample collection [3], and some space robots have been used in space exploration [4]. Although ground mobile robots, underwater robots, and space robots have been widely used, the application of aerial

robots is still in its infancy all over the world. The current unmanned aerial vehicles play an important role in monitoring activities, such as aerial photography and high-voltage line inspection [5], but do not have the ability to physically interact with the environment. However, the aerial operation robot (aerial manipulator) has great application value in the following three aspects: (1) replacing manpower to complete dangerous tasks, such as urban antiterrorism and high-rise building firefighting [6]; (2) replacing manpower to improve efficiency, as in wide-area scientific examination and collection [7]; (3) replacing manpower to reduce costs, as in infrastructure maintenance and remote operation in complex environment [8].

The configuration of the aerial robot commonly consists of an aerial platform and an operation manipulator, but their combination brings completely new features to the system. The main challenges in the aerial robot system design include two aspects. First, since the aerial robot usually works



in confined spaces that require operation, it should have high traffic ability and contact ability with the complex environment. Under this premise, it also should have as large operating payload as possible. Second, there is a serious coupling effect between the aerial platform and the robot arm, which makes the system face enormous challenge in terms of stability and manipulation accuracy, and physical contact and manipulation further exacerbate it. The research of aerial robots has begun to attract worldwide attention since 2010, and the representative work includes AIRobots Project [9], ARCAS Project [10], AEROWORKS Project [11].

In respect to structure, most of aerial robots use helicopter or multirotor as aerial platform. Yale University designed a helicopter with a single DOF (degree of freedom) underactuated gripper, as shown in Figure 1(a), and studied the quasi-static compliance control problem during the grabbing process through PID control method [12]. University of Drexel also used helicopter as base platform to expand the 1-DOF gripper into a multi-DOF manipulator and completed the grasping and placement tasks of cylindrical objects [13]. National Taipei University of Technology designed a small quadrotor with a 2-DOF manipulator and carried out simulation analysis of kinematic control in 2D plane based on a simplified vector model [14]. University of Seville designed a dual-arm system based on quadrotor, as shown in Figure 1(b), which expands the maneuvering range and enables more complex operations [15]. However, because of the inherent characteristics of open rotor, the helicopter or multirotor structure is unable to interact with the environment closely, and the arm can only manipulate objects above [10] or below [12], which greatly limits the application scenarios. In order to grab the target side-on, Johns Hopkins University proposed a very long arm to avoid the rotor disc of the quadrotor platform [16], but the changes in the center of mass and moment of inertia caused by the long arm exert serious impacts on the stability and effective payload of the system. Compared to helicopter and multirotor, ducted fan has greater thrust in a more compact structure [17], having contact capability and increasing the payload of the system, which is more suitable as the platform of an aerial robot. Figure 2 shows the comparison of flight area requirements of different types of aerial robots near a wall (namely, our innovative ducted fan aerial robot, quadrotor, and helicopter). It can be seen that, under the same effective load, ducted fan aerial robot can be closer to the target and operate it from side-on with a smaller joint motion range and can pass narrower confined space and be safer. University of Bologna completed a series of studies on contact dynamics of the single ducted fan with vertical wall [18], but their platform cannot carry the manipulator due to the payload and controllability limitations.

With regard to control, because of the serious coupling between aerial platform and manipulator, on the one hand, the manipulation process will have three effects on the aerial platform: (1) the effects of the manipulator gravity, and the gravity moment caused by the noncoincidence of the center of mass between the aerial vehicle and the manipulator, (2) the inertia force and moment generated

by manipulator dynamics, and (3) the impact of the external environment during contacting and operating; on the other hand, the drift of the aerial platform also affects the positioning accuracy of the end effector [19]. Some studies ignored the existence of manipulator and only considered the aerial vehicle [20], and some tried to improve the robustness of the aerial vehicle basic controller to ensure system stability [21–23], but both of them are only suitable for small size arm (e.g., lightweight 1-DOF gripper) and lightweight object. DLR in [20] proposed an impedance controller to stabilize the system in the presence of external forces, ignoring the coupling effects between the aircraft and the arm by limiting the motion of the arm. Shenyang Institute of Automation of CAS in [23] designed a linear LQR controller of a helicopter with 1-DOF arm. The simulation results show that the controller has good performance when the arm moves in a small range near the equilibrium point, but the LQR controller cannot stabilize the system when the swing range of the arm is relatively large. AGH University of Science and Technology considered the change in the system's center of mass caused by the arm movement and compensated for the influence through variable-parameter PID control to achieve system's stability [24], while the University of Pennsylvania considered the impact of the payload acting on the system's center of mass [25]. The University of Seville designed a controller to compensate for the change of both the center of mass and the moment of inertia when the arm is in different positions, but the dynamic characteristics of the arm are not considered [26]. The University of Naples Federico II designed a Cartesian space impedance controller based on the integrated dynamic model of the aerial manipulator, which fully considered the coupling effect between the aerial vehicle and the robot arm [27]. However, the controller is highly dependent on the model and has a complicated structure, which is not easy in practice.

Based on the discussion above, the contribution of this paper mainly includes two parts. First, an innovative aerial manipulator based on tandem ducted fans is proposed, which has both great trafficability and effective payload. The dynamic model of the aerial manipulator is established by comprehensive mechanism and parametric frequency domain identification. Owing to the small lateral size, the aerial manipulator can easily realize the omnidirectional manipulation of side-on and below and is particularly suitable for the tasks in confined environment, where traditional multirotor and helicopter would be inaccessible. The application aims of this novel design are canopy sampling in dense forests and insulator lubricating in dense high-voltage wires. Second, a composite controller of the aerial platform is proposed, considering and compensating for both the static and the dynamic disturbances of the manipulator on the vehicle. A basic static robust controller is designed via H-infinity synthesis for basic performance, and an additional adaptive loop is designed for disturbance estimation and compensation from the manipulator to improve the platform stability and the end effector tracking accuracy. The computer simulations verify the effectiveness of the proposed controller.



FIGURE 1: Examples of aerial robots. (a) Yale University. (b) University of Seville.

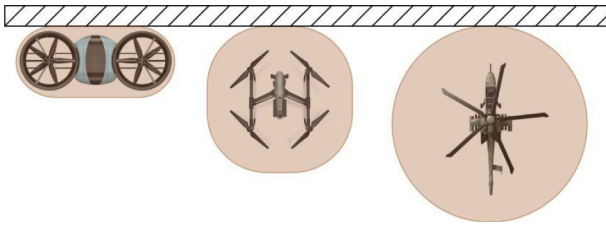


FIGURE 2: Comparison of flight area requirements of different types of aerial robots.

The organization of this paper is as follows. In Section 2, the structure of the innovative ducted fan aerial manipulator is introduced and the dynamic model is established. Then, based on the dynamic model, a control-oriented state space model of the aerial platform is derived based on parametric identification to facilitate the controller design. In Section 3, the basic robust controller and the adaptive auxiliary controller are designed and analyzed in detail. In Section 4, the simulation results and analysis of the proposed controller are carried out. Finally, some conclusions are drawn in Section 5.

## 2. Modeling of the Ducted Fan Aerial Manipulator

**2.1. System Description.** The configuration of the proposed aerial manipulator mainly includes two ducted fan systems with coaxial rotors, two sets of control rudders, a 3-DOF manipulator, the control unit, and the landing gears, as shown in Figure 3. The pitch channel of the vehicle is controlled by the thrust difference between the front and the rear ducted fans, which is caused by the difference of rotor speed; the roll channel is controlled by the rudder systems setting below the duct; the yaw channel is controlled by the reaction torque difference between the upper and lower rotor disc. The total mass of the aerial vehicle is 4.6 kg, with an effective payload of 2 kg. The manipulator weight is 0.9 kg, with a max grasping weight of 0.5 kg. The parameters of the system are detailed in Table 1, where the mass and the structural dimension parameters of the platform are measured directly, and the moments of inertia and the effective

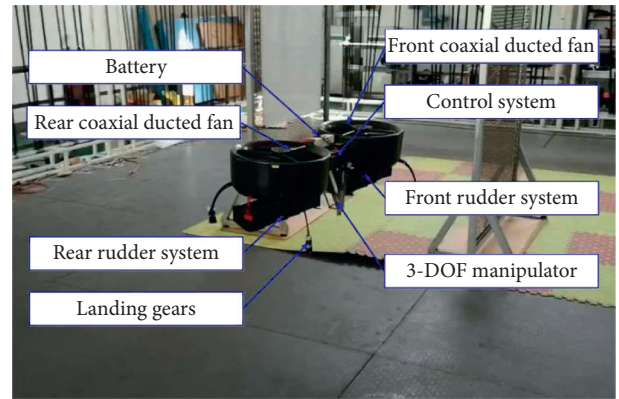


FIGURE 3: The configuration of the novel aerial manipulator.

TABLE 1: Structural parameters of the system.

| Parameter     | Physical description                               | Value                   |
|---------------|--|-------------------------|
| $m_b$         | Vehicle mass (including battery)                   | 4.6 kg                  |
| $m_r$         | Manipulator mass                                   | 0.9 kg                  |
| $m_{grasp}$   | Max grasping mass                                  | 0.5 kg                  |
| $I_{xx}$      | Inertia tensor of the vehicle around $x$ -axis     | 0.092 kg·m <sup>2</sup> |
| $I_{yy}$      | Inertia tensor of the vehicle around $y$ -axis     | 0.283 kg·m <sup>2</sup> |
| $I_{zz}$      | Inertia tensor of the vehicle around $z$ -axis     | 0.245 kg·m <sup>2</sup> |
| $p_{cd}$      | Distance between duct center and CG of vehicle     | 0.32 m                  |
| $D$           | Duct diameter                                      | 0.33 m                  |
| $S_x$         | Vehicle effective resistance area around $x$ -axis | 0.034 m <sup>2</sup>    |
| $S_y$         | Vehicle effective resistance area around $y$ -axis | 0.127 m <sup>2</sup>    |
| $S_z$         | Vehicle effective resistance area around $z$ -axis | 0.133 m <sup>2</sup>    |
| $n$           | Blade number of each disc                          | 4                       |
| $c$           | Blade chord length                                 | 0.027 m                 |
| $\theta_0$    | Attack angle at the root of blade                  | 35 deg                  |
| $\theta_{rw}$ | Torsion rate of blade                              | 18 deg                  |
| $\sigma$      | Blade tip clearance                                | 0.001 m                 |

resistance areas along three axes of the vehicle are estimated by the CATIA® 3D model. The airfoil used in the prototype is NACA 0012, and the airfoil parameters can be obtained from [28]. The manipulator is installed on the center of the vehicle body and has 3 DOF (one lumbar joint, one shoulder joint, and one elbow joint) to reach any position in 3D space. The end effector is a gripper, and, from the perspective of our application purpose, the posture of the gripper is not considered. The parameters of the arm are described using standard D-H (Denavit–Hartenberg) method [29], as shown in Table 2.

The hardware of control unit is shown in Figure 4. The on-board controller is Emlid® Navio2 based on Raspberry Pi® 3, which integrates dual IMU module, GPS module, barometer module, and 14 PWM output channels. The rotors are driven by four GARTT® motors, and the motors are driven by HOBBYWING® Electronic Speed Controllers (ESC), which are controlled by the main controller via PWM signals. Each duct system has two sets of control rudders, which are driven synchronously by one KST® servo. The servos are also controlled by PWM signals from the main controller. The Dynamixel® XH430 servo is chosen as arm joint servo, since it has both position control mode and torque control mode and has feedback function of actual position, velocity, and torque of the joint. The on-board main controller communicates with the ground station through the 3DR® radio telemetry.

**2.2. Dynamic Model.** The coordinate system of the aerial manipulator is introduced in Figure 5. The aerial manipulator is modeled as a multibody system consisting of four interconnected rigid bodies. Let  $\Sigma$  be the earth-fixed Cartesian coordinate frame following the north-east-down rules,  $\Sigma_b$  be the vehicle body-fixed coordinate frame at the center of mass of the vehicle,  $\Sigma_0$  be the manipulator base-fixed coordinate frame, and  $\Sigma_i$  ( $i = 1, 2, 3$ ) be the coordinate frame of each link of the manipulator following the D-H rules. Notice that  $\Sigma_0$  coincides with the origin of  $\Sigma_b$ , only rotated  $90^\circ$  around the  $Z_b$  axis. In the following, the superscript  $i$  means that the variable is related to the coordinate system  $\Sigma_i$ .

The dynamic formulation is established using iterative Newton–Euler method [29], which is widely used in dynamic modeling [30, 31]. First, give the definition of the motion states of the system. Let  $p_b = [x \cdot y \cdot z]^T$  and  $\Phi_b = [\varphi \cdot \theta \cdot \psi]^T$  be the position vector and Euler angle vector of the aerial vehicle platform in earth-fixed frame, and let  $q_i \cdot (i = 1 \cdot 2 \cdot 3)$  be the joint angle of the manipulator. Let  $v_b$  and  $\omega_b$  be the velocity vector and angular rate vector of the aerial vehicle, and let  $v_i$  and  $\omega_i$  be the velocity and angular rate of the origin of link  $i$  of the manipulator. Then, the acceleration of the multibody system can be calculated outward iteratively from the aerial vehicle platform to the manipulator end effector as

TABLE 2: D-H parameters of the manipulator.

| Link | a (m) | $\alpha$ (deg) | d (m) | $\theta$ range (deg) |
|------|-------|----------------|-------|----------------------|
| 1    | 0     | 90             | 0.08  | 180 $[-90, 90]$      |
| 2    | 0.15  | 0              | 0     | 150 $[0, 150]$       |
| 3    | 0.16  | 0              | 0     | 150 $[-150, 0]$      |

$$\left\{ \begin{array}{l} \mathbf{v}_b^b = [u \ v \ w]^T = \dot{p}_b^b = \mathbf{R}_b^T \dot{p}_b, \\ \boldsymbol{\omega}_b^b = [p \ q \ r]^T = \mathbf{R}_b^T \boldsymbol{\omega}_b = \mathbf{R}_b^T \mathbf{T}_b \dot{\Phi}_b = \mathbf{Q}_b \dot{\Phi}_b, \\ \mathbf{v}_0^0 = \mathbf{R}_b^0 \mathbf{v}_b^b, \\ \boldsymbol{\omega}_0^0 = \mathbf{R}_b^0 \boldsymbol{\omega}_b^b, \\ \boldsymbol{\omega}_i^i = \mathbf{R}_{i-1}^i \boldsymbol{\omega}_{i-1}^{i-1} + \dot{q}_i \mathbf{e}_{zi}, \\ \dot{\boldsymbol{\omega}}_i^i = \mathbf{R}_{i-1}^i \dot{\boldsymbol{\omega}}_{i-1}^{i-1} + \mathbf{R}_{i-1}^i \boldsymbol{\omega}_{i-1}^{i-1} \times \dot{q}_i \mathbf{e}_{zi} + \ddot{q}_i \mathbf{e}_{zi}, \\ \dot{v}_i^i = \mathbf{R}_{i-1}^i (\dot{\boldsymbol{\omega}}_{i-1}^{i-1} \times \mathbf{p}_i^{i-1} + \boldsymbol{\omega}_{i-1}^{i-1} \times (\boldsymbol{\omega}_{i-1}^{i-1} \times \mathbf{p}_i^{i-1})) + \dot{v}_{i-1}^{i-1}, \\ \mathbf{a}_{C_i}^i = \dot{\boldsymbol{\omega}}_i^i \times \mathbf{p}_{C_i}^i + \boldsymbol{\omega}_i^i \times (\boldsymbol{\omega}_i^i \times \mathbf{p}_{C_i}^i) + \dot{v}_i^i, \end{array} \right. \quad (1)$$

where the superscript  $b$  refers to the body-fixed coordinate system, the superscript  $i$  refers to the  $i^{\text{th}}$  link coordinate system, and in particular  $i = 0$  for the manipulator base-fixed (link 0) coordinate system, as shown in Figure 5. Specifically,  $v_b^b$  and  $\omega_b^b$  are the velocity and angular rate of the aerial vehicle with respect to the body-fixed coordinate frame,  $v_i^i$  and  $\omega_i^i$  are the velocity and angular rate of the origin of link  $i$  with respect to the  $i^{\text{th}}$  link coordinate frame.  $\mathbf{a}_{C_i}^i$  is the acceleration of the center of mass of link  $i$ .  $\mathbf{p}_i^{i-1}$  is the position vector of the origin of  $i^{\text{th}}$  link frame with respect to the  $(i-1)^{\text{th}}$  link coordinate frame,  $\mathbf{p}_{C_i}^i$  is the position vector of the center of mass of  $i^{\text{th}}$  link with respect to the  $i^{\text{th}}$  link coordinate frame, and  $\mathbf{e}_{zi}$  refers to the projection along the  $z$ -axis of link  $i$ .  $\mathbf{R}_{i-1}^i$  refers to the transformation matrix from the  $(i-1)^{\text{th}}$  link coordinate frame to the  $i^{\text{th}}$  link coordinate frame. In particular,  $\mathbf{R}_b^0$  refers to the transformation matrix from the body-fixed frame to the arm base-fixed (link 0) frame.  $\mathbf{R}_b$  and  $\mathbf{Q}_b$  are the linear velocity and angular rate transformation matrix between the earth-fixed frame and body-fixed frame expressed in the form of Euler angles.

Then, the Newton–Euler dynamics formulation can be derived as

$$\left\{ \begin{array}{l} \mathbf{F}_b^b = m_b (\dot{v}_b^b + \boldsymbol{\omega}_b^b \times v_b^b), \\ \mathbf{M}_b^b = \mathbf{I}_b \dot{\boldsymbol{\omega}}_b^b + \boldsymbol{\omega}_b^b \times \mathbf{I}_b \boldsymbol{\omega}_b^b, \\ \mathbf{F}_i^i = m_i \mathbf{a}_{C_i}^i, \\ \mathbf{M}_i^i = \mathbf{I}_i \dot{\boldsymbol{\omega}}_i^i + \boldsymbol{\omega}_i^i \times \mathbf{I}_i \boldsymbol{\omega}_i^i, \end{array} \right. \quad (2)$$

where  $m_b$  and  $m_i$  are the mass of the aerial vehicle and link  $i$ ,  $\mathbf{I}_b$  and  $\mathbf{I}_i$  are the inertia matrix of the aerial vehicle and link  $i$ .  $\mathbf{F}_b^b$  and  $\mathbf{F}_i^i$  are the total external force exerted on the aerial vehicle and link  $i$ , and  $\mathbf{M}_b^b$  and  $\mathbf{M}_i^i$  are the total external moment exerted on the aerial vehicle and link  $i$ , which can be calculated inward iteratively from the manipulator end effector to the aerial vehicle platform as

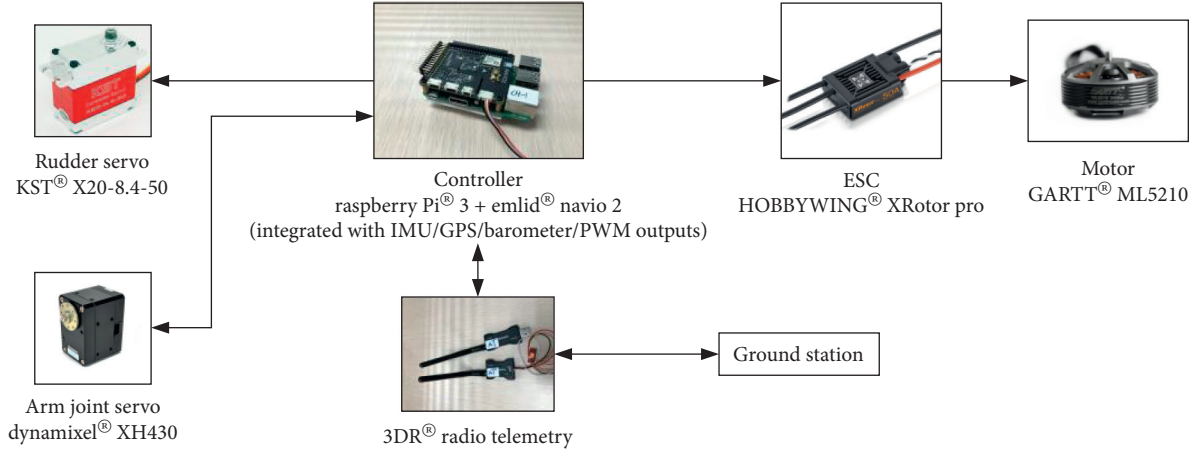


FIGURE 4: The hardware of control unit.

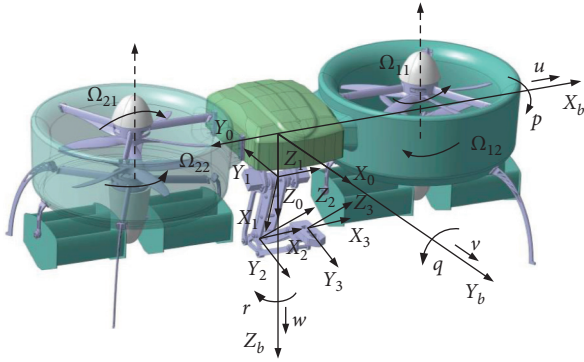


FIGURE 5: The coordinate system of the aerial manipulator.

$$\begin{cases}
 \mathbf{f}_i^i = \mathbf{F}_i^i + \mathbf{R}_{i+1}^i \mathbf{f}_{i+1}^{i+1} - \mathbf{R}_i^T m_i \mathbf{g}, \\
 \mathbf{n}_i^i = \mathbf{M}_i^i + \mathbf{R}_{i+1}^i \mathbf{n}_{i+1}^{i+1} + \mathbf{p}_{i+1}^i \times \mathbf{R}_{i+1}^i \mathbf{f}_{i+1}^{i+1} + \mathbf{p}_{C_i}^i \times (\mathbf{F}_i^i - \mathbf{R}_i^T m_i \mathbf{g}), \\
 \tau_i = (\mathbf{n}_i^i)^T \mathbf{e}_{z_i}, \\
 \mathbf{F}_b^b = \mathbf{F}_{\text{aero1}} + \mathbf{F}_{\text{aero2}} + \mathbf{F}_{\text{rudder1}} + \mathbf{F}_{\text{rudder2}} + \mathbf{F}_{\text{fus}} + \mathbf{R}_b^T m_b \mathbf{g} \\
 \quad - \mathbf{R}_0^b \mathbf{R}_1^0 \mathbf{f}_1^1, \\
 \mathbf{M}_b^b = \mathbf{M}_{\text{aero1}} + \mathbf{M}_{\text{aero2}} + \mathbf{p}_{cd1}^b \times \mathbf{F}_{\text{aero1}} + \mathbf{p}_{cd2}^b \times \mathbf{F}_{\text{aero2}} + \mathbf{p}_{cr1}^b \\
 \quad \times \mathbf{F}_{\text{rudder1}} + \mathbf{p}_{cr2}^b \times \mathbf{F}_{\text{rudder2}} + \mathbf{M}_{\text{gyro}} - \mathbf{R}_0^b \mathbf{R}_1^0 \mathbf{n}_1^1,
 \end{cases} \quad (3)$$

where  $\mathbf{f}_i^i$  and  $\mathbf{n}_i^i$  are the force and moment exerted on link  $i$  by link  $i-1$  and  $\mathbf{R}_i$  is the transformation matrix from the  $i^{\text{th}}$  link coordinate frame to the earth-fixed coordinate frame.  $\tau_i$  is the joint torque of the manipulator.

The total force of the aerial vehicle includes five parts: the aerodynamic force of the ducted fans, the resultant force generated by the rudders, the fuselage resistance, the gravity, and the reaction force of the manipulator. The total moment of the aerial vehicle also includes five parts: the aerodynamic moment of the ducted fans, the torque generated by the aerodynamic forces, the torque generated by the rudders, the reaction torque of the manipulator, and the gyro moment

produced by rotors, as shown in Figure 6. In equation (3), the subscript ‘‘aero’’ refers to the aerodynamic force and moment, subscript 1 denotes the front ducted fan system, and subscript 2 denotes the rear ducted fan system. Same as the above, the subscript rudder refers to the force of the front and the rear rudder system. fus refers to the fuselage resistance, and gyro refers to the gyro torque.  $\mathbf{p}_{cd}^b$  is the position vector of the center of duct with respect to the center of the vehicle, and  $\mathbf{p}_{cd}^b$  is the position vector of the aerodynamic center of rudder with respect to the center of the vehicle. These forces and moments belong to the ducted fan system dynamics, which provide the main thrust and attitude control moments for the aerial vehicle, and will be introduced in detail in the following section.

**2.3. Ducted Fan System.** As mentioned in the Introduction, most of the literature of aerial manipulator uses multirotor as aerial platform. Because of the simple, symmetrical, and decoupled open rotor structure, the rotor aerodynamics is usually simplified as a thrust and a reaction torque in quadratic relation to the rotor speed, ignoring any other effects of the aerodynamic characteristics (e.g., [22, 26, 27]). However, the aerodynamics of ducted fan is significantly different from the traditional open rotor [32]. First, the duct lip causes airflow deflection effect, which changes the inflow at the rotor disc, while additional thrust is generated by the suction flow of the duct lip; second, duct suppresses the rotor tip vortex and reduces the momentum loss; third, the duct exit prevents the airflow from contracting, which increases the outflow pressure of the rotor and thereby increases the thrust. Johnson and Turbe [17] propose a Blade Element Momentum Theory (BEMT) model for ducted-single-rotor system. By improving this method, the dynamic model of our ducted-coaxial-rotor system is proposed. The basic idea of BEMT is to establish a set of relationships between rotor aerodynamic force and induced velocity through momentum theorem and blade element theory, respectively, and then solve the simultaneous equations iteratively. The following is the derivation of the front ducted fan, and the rear one can be obtained in the same way.

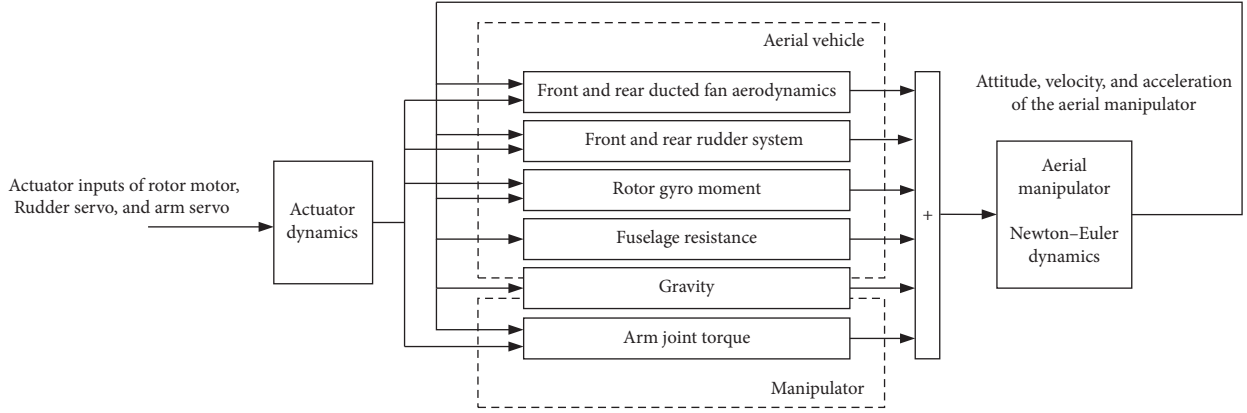


FIGURE 6: Diagram of the dynamic model.

*Momentum Theorem.* The inflow process of the coaxial ducted fan under airflow deflection effect is shown in Figure 7. It can be divided into five parts: the free inflow, the deflected airflow whose direction is deflected by duct lip, the airflow at upper rotor disc, the airflow at lower rotor disc, and the outflow of duct exit. The free flow at the ducted fan is

$$\mathbf{V}_a = [u_a \ v_a \ w_a]^T = \mathbf{v}_b^b - \mathbf{R}_b^T \mathbf{v}_w, \quad (4)$$

where  $\mathbf{v}_b^b$  is the body velocity of the aerial vehicle and  $\mathbf{v}_w$  is the local wind velocity. Express  $v_a$  in the inflow plane; the inflow velocity  $v_0$  is calculated as

$$\begin{cases} \mathbf{V}_0 = |\mathbf{V}_0| \cos \alpha i + |\mathbf{V}_0| \sin \alpha j, \\ |\mathbf{V}_0| = |\mathbf{V}_a|, \alpha = \arctan\left(\frac{w_a}{\sqrt{u_a^2 + v_a^2}}\right). \end{cases} \quad (5)$$

Then, the deflected airflow, the airflow at upper and lower rotor disc, and the outflow can be calculated by

$$\begin{cases} \mathbf{V}_t = |\mathbf{V}_0| \cos \alpha_t i + |\mathbf{V}_0| \sin \alpha_t j, \\ \mathbf{V}_{r1} = |\mathbf{V}_0| \cos \alpha_t i + (|\mathbf{V}_0| \sin \alpha_t + v_{i1}) j, \\ \mathbf{V}_{r2} = |\mathbf{V}_0| \cos \alpha_t i + (|\mathbf{V}_0| \sin \alpha_t + v_{i1} + v_{i2}) j, \\ \mathbf{V}_e = |\mathbf{V}_0| \cos \alpha_t i + (|\mathbf{V}_0| \sin \alpha_t + v_{\infty}) j, \end{cases} \quad (6)$$

where  $v_{i1}$  and  $v_{i2}$  refer to the induced velocities of upper and lower rotor,  $v_{\infty}$  refers to the overall impact of the ducted fan on free flow, and  $\alpha_t$  is the deflection angle caused by duct effect at duct lip, which can be expressed as a function of free inflow angle and airflow deflection factor as

$$\alpha_t = \alpha + k_t \left( \frac{\pi}{2} - \alpha \right). \quad (7)$$

According to momentum theorem and kinetic energy theorem, considering the control body from the free inflow

to the upper disc exit and from the upper disc exit to the duct exit and the entire duct, respectively, the following equations are obtained:

$$\begin{cases} T_{r1} (|\mathbf{V}_0| \sin \alpha_t + v_{i1}) = \frac{1}{2} \rho S_D |\mathbf{V}_0| \left( (|\mathbf{V}_0| \sin \alpha_t + v_{i1})^2 - (|\mathbf{V}_0| \sin \alpha)^2 \right), \\ T_{r2} (|\mathbf{V}_0| \sin \alpha_t + v_{i1} + v_{i2}) = \frac{1}{2} \rho S_D |\mathbf{V}_0| \left( (|\mathbf{V}_0| \sin \alpha_t + v_{\infty})^2 - (|\mathbf{V}_0| \sin \alpha_t + v_{i1})^2 \right), \\ T = (1 + k_{\text{duct}}) (T_{r1} + T_{r2}) = \rho S_D |\mathbf{V}_0| \left( (|\mathbf{V}_0| \sin \alpha_t + v_{\infty})^2 - (|\mathbf{V}_0| \sin \alpha)^2 \right), \\ T_d = \rho S_D |\mathbf{V}_0| (|\mathbf{V}_0| \cos \alpha_t - |\mathbf{V}_0| \cos \alpha), \end{cases} \quad (8)$$

where  $T_{r1}$  and  $T_{r2}$  are the thrusts of upper rotor and lower rotor in the vertical direction,  $T_d$  is the momentum resistance in the horizontal plane,  $k_{\text{duct}}$  is the thrust augment factor of duct, and  $SD$  is the rotor disc area.

*Blade Element Theory.* Blade element theory analyzes the aerodynamics of each micro element of the blade, then integrates it along the radial direction of the blade, and averages it along the circumferential direction to solve the forces and moments of the entire rotor. According to equations (5) and (6), the airflow velocities at the blades of upper and lower rotors are

$$\begin{cases} \mathbf{V}_{ra1} = [u_{ra1} \ v_{ra1} \ w_{ra1}]^T = \left[ |\mathbf{V}_0| \cos \alpha_t \frac{u_a}{\sqrt{u_a^2 + v_a^2}} \quad |\mathbf{V}_0| \cos \alpha_t \frac{v_a}{\sqrt{u_a^2 + v_a^2}} \quad |\mathbf{V}_0| \sin \alpha_t \right]^T, \\ \mathbf{V}_{ra2} = [u_{ra2} \ v_{ra2} \ w_{ra2}]^T = \left[ |\mathbf{V}_0| \cos \alpha_t \frac{u_a}{\sqrt{u_a^2 + v_a^2}} \quad |\mathbf{V}_0| \cos \alpha_t \frac{v_a}{\sqrt{u_a^2 + v_a^2}} \quad |\mathbf{V}_0| \sin \alpha_t + v_{i1} \right]^T. \end{cases} \quad (9)$$

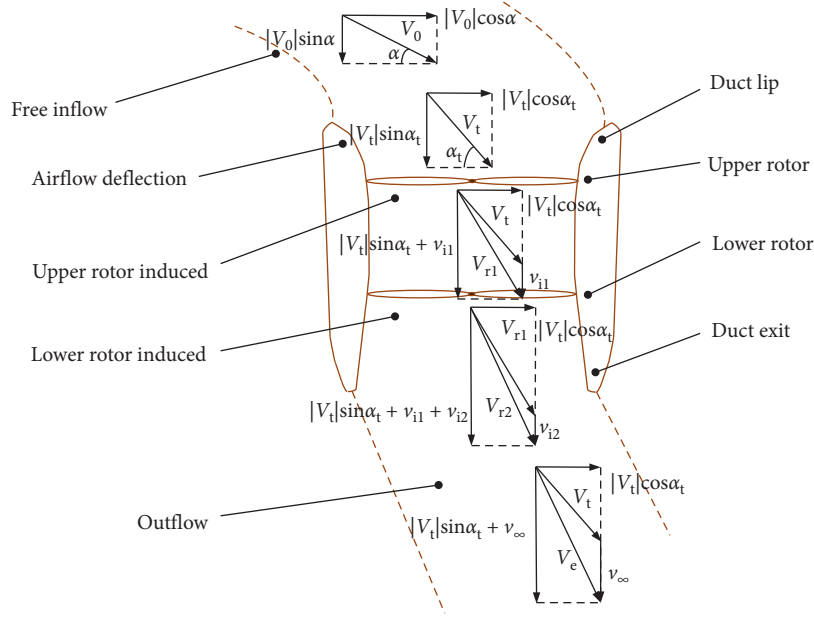


FIGURE 7: The inflow process model.

Since the blade element theory is a common method, the results after integration of the rotor thrust  $T_{ri}$ , the reaction torque  $Q_{ri}$ , the pneumatic rolling torque  $L_{ri}$ , and the

pneumatic pitching torque  $M_{ri}$  of the upper and lower rotor are directly given below. Notice that  $i = 1$  for upper rotor and  $i = 2$  for lower rotor:

$$\left\{ \begin{array}{l} T_{ri} = \frac{n}{2\pi} \int_0^{2\pi} \int_0^R dT_{ri} d\psi_0 = \frac{n\rho c C_a \Omega_i^2 R^3}{4} \left( \left( \frac{2}{3} + \mu_{ui}^2 + \mu_{vi}^2 \right) \theta_0 - (1 + \mu_{ui}^2 + \mu_{vi}^2) \frac{\theta_{rw}}{2} - \lambda_{wi} \right), \\ Q_{ri} = \frac{n}{2\pi} \int_0^{2\pi} \int_0^R dQ_{ri} d\psi_0 = \frac{n\rho c C_a \Omega_i^2 R^4}{4} \left( \left( \frac{2}{3} \theta_0 - \frac{1}{2} \theta_{rw} - \lambda_{wi} \right) \lambda_{wi} + \frac{1}{2} (1 + \mu_{ui}^2 + \mu_{vi}^2) C_d \right), \\ L_{ri} = \frac{n}{2\pi} \int_0^{2\pi} \int_0^R dL_{ri} d\psi_0 = \frac{n\rho c C_a \Omega_i^2 R^4}{4} \left( \mu_{vi} \left( \frac{2}{3} \theta_0 - \frac{1}{2} \lambda_{wi} - \frac{1}{2} \theta_{rw} \right) \right), \\ M_{ri} = \frac{n}{2\pi} \int_0^{2\pi} \int_0^R dM_{ri} d\psi_0 = \frac{n\rho c C_a \Omega_i^2 R^4}{4} \left( \mu_{ui} \left( \frac{2}{3} \theta_0 - \frac{1}{2} \lambda_{wi} - \frac{1}{2} \theta_{rw} \right) \right). \end{array} \right. \quad (10)$$

In the above formula,  $n$  is the blade number of each disc,  $R$  is the rotor radius,  $\psi_0$  is the blade azimuth angle,  $c$  is the blade chord length,  $C_a$  is the lift coefficient slope,  $C_d$  is the resistance coefficient,  $\Omega_i$  is the rotor speed,  $\theta_0$  is the blade attack angle, and  $\theta_{rw}$  is the blade torsion rate.  $\mu_{ui}$ ,  $\mu_{vi}$ , and  $\lambda_{wi}$  denote the dimensionless ratios of forward velocity, lateral velocity, and inflow velocity:

$$\mu_{ui} = \left( \frac{u_{rai}}{\Omega_i R} \right), \mu_{vi} = \left( \frac{v_{rai}}{\Omega_i R} \right), \lambda_{wi} = \left( \frac{v_{ii} - w_{rai}}{\Omega_i R} \right). \quad (11)$$

Solve equations (8) and (10) simultaneously; the aero force and moment of the ducted fan in body-fixed frame can be derived as

$$\left\{ \begin{array}{l} \mathbf{F}_{aero} = \left[ -\frac{u_a}{\sqrt{u_a^2 + v_a^2}} T_d \quad -\frac{v_a}{\sqrt{u_a^2 + v_a^2}} T_d \quad -T \right]^T, \\ \mathbf{M}_{aero} = [L_{r1} - L_{r2} \quad M_{r1} - M_{r2} \quad Q_{r1} - Q_{r2}]^T. \end{array} \right. \quad (12)$$

*Rudder Dynamics.* Same as the ducted fan system, the following gives the analysis of the front rudder, and the rear one can be obtained in the same way. The rudder is in the outflow of the duct exit, and its attack angle can be calculated by

$$\alpha_v = \alpha_e + \delta, \quad (13)$$

where  $\alpha_e$  is the angle of the duct outflow velocity  $V_e$ , and  $\delta$  is the rudder control angle. Then the aero-lift and aero-drag of the rudder can be derived as

$$\begin{cases} L_v = \left(\frac{1}{2}\right)\rho|\mathbf{V}_e|^2 S_v C_{lv}, \\ D_v = \left(\frac{1}{2}\right)\rho|\mathbf{V}_e|^2 S_v C_{dv}, \end{cases} \quad (14)$$

where  $\rho$  is the airflow density,  $S_v$  is the rudder area, and  $C_{lv}$  and  $C_{dv}$  are the lift coefficient and drag coefficient, which are the functions of  $\alpha_v$ . The force of the control rudder in body-fixed frame can then be derived.  $n_v$  is the number of the rudders in one duct:

$$\mathbf{F}_{\text{rudder}} = [0 \quad n_v(L_v \cos \alpha_e + D_v \sin \alpha_e) \quad n_v(-L_v \sin \alpha_e + D_v \cos \alpha_e)]^T. \quad (15)$$

In summary, the expression of each component in equation (3) is obtained, and the comprehensive nonlinear mechanism model is established.

**2.4. Model Identification of the Aerial Platform.** The comprehensive nonlinear mechanism model established in the previous section can fully reflect the characteristics of the system in all-envelope range and be used for system performance prediction and comprehensive simulation analysis, but the complex mechanism is not suitable for the controller design process [33]. Since the working scenarios of the aerial vehicle in this paper are hovering or near-hovering conditions, in order to reduce the complexity of nonlinear identification and improve the identification accuracy under the noise of low-cost sensors, the parametric frequency domain identification is performed. First, a simplified parametric model can be derived from the nonlinear mechanism model of the aerial platform by small perturbation assumption under hover equilibrium point; then, a control-oriented state space model of the aerial platform is derived based on frequency domain identification method. The identification model is verified in both time and frequency domains.

The nonlinear mechanism model of the aerial vehicle can be described as

$$\begin{cases} \dot{\mathbf{x}}_b = \mathbf{f}(\mathbf{x}_b, \mathbf{u}_b), \\ \mathbf{x}_b = [u \quad v \quad w \quad p \quad q \quad r \quad \varphi \quad \theta \quad \psi]^T, \\ \mathbf{u}_b = [u_{\text{alt}} \quad u_{\text{rol}} \quad u_{\text{pit}} \quad u_{\text{yaw}}]^T, \end{cases} \quad (16)$$

where  $\mathbf{u}_b$  is the normalization control input vector, referring to the altitude channel, roll channel, pitch channel, and yaw channel, and is mapped to the rotor speed difference and rudder angle (described in 2.1). Find the partial derivative of equation (16) at hover equilibrium point to obtain the system's Jacobian matrix:

$$\begin{cases} \dot{\mathbf{x}}_b \approx \mathbf{A}(\mathbf{x}_b - \mathbf{x}_{\text{hover}}) + \mathbf{B}(\mathbf{u}_b - \mathbf{u}_{\text{hover}}), \\ \mathbf{A} = \left(\frac{\partial \mathbf{f}}{\partial \mathbf{x}_b}\right), \mathbf{B} = \left(\frac{\partial \mathbf{f}}{\partial \mathbf{u}_b}\right), \\ \mathbf{f}(\mathbf{x}_{\text{hover}}, \mathbf{u}_{\text{hover}}) = \mathbf{0}. \end{cases} \quad (17)$$

Then, the simplified parametric matrix A and B to be identified is obtained by small perturbation assumption. Since  $\mathbf{x}_{\text{hover}}$  and  $\mathbf{u}_{\text{hover}}$  are constant and  $\mathbf{x}_{\text{hover}} = \mathbf{0}$ , in the rest of this paper, for the simplicity of the symbol, let  $\mathbf{x}_b \equiv \mathbf{x}_b - \mathbf{x}_{\text{hover}}$ ,  $\mathbf{u}_b \equiv \mathbf{u}_b - \mathbf{u}_{\text{hover}}$ . Because the aerial vehicle platform is inherently unstable, the closed-loop identification is carried out following the procedure in [34]. A simple PID controller is used to ensure the stability during identification process. A set of Chirp sweep signals (introduced in [35]) is performed in each of the input channels, and the data is processed by CIFER<sup>®</sup> identification tool [35]. The model fitting process is an optimization problem as

$$\min J = \sum_{i=1}^n \left\{ \left( \frac{20}{n_\omega} \right) \sum_{\omega_1}^{\omega_{n_\omega}} W_\gamma \left[ W_g \left( |G_{fi}| - |G_i| \right)^2 + W_p \left( \angle G_{fi} - \angle G_i \right)^2 \right] \right\}, \quad (18)$$

where  $G_{fi}$  is the transfer function to be fitted corresponding to  $G_i$ , converting from the parametric state space equation, and  $n$  is the number of the transfer functions.  $|G|$  refers to the amplitude and  $\angle G$  refers to the phase.  $n_\omega$  is the number of the selected frequency sampling points, and  $\omega_1$  and  $\omega_{n_\omega}$  refer to the initial and termination sampling frequency respectively.  $W_\gamma$ ,  $W_g$ , and  $W_p$  are the weighting functions referring to coherence value, amplitude, and phase, respectively, and are chosen following the directions in CIFER<sup>®</sup>. The frequency response identification results are shown in Figures 8 and 9, and the identified model is

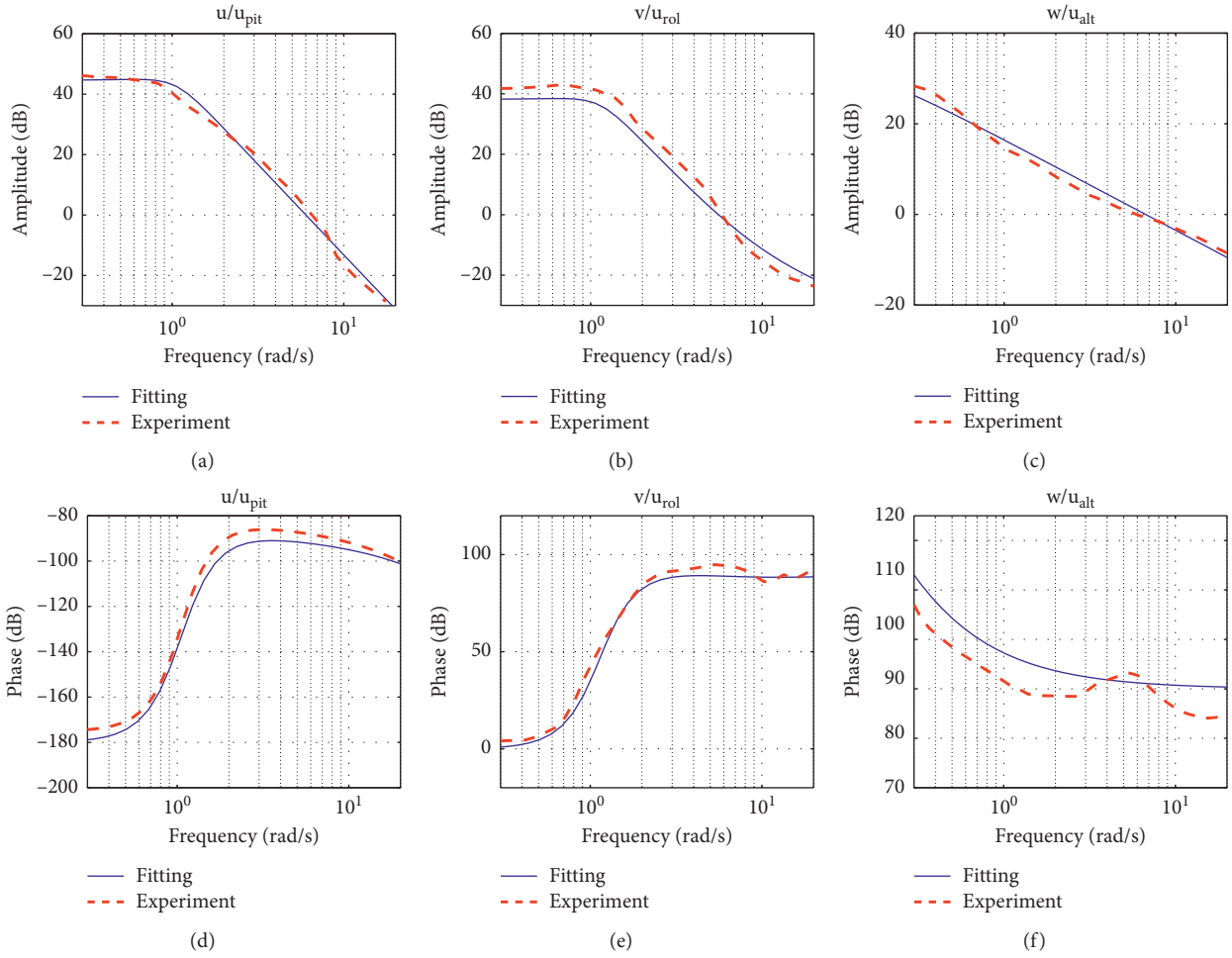


FIGURE 8: Frequency response identification results of linear velocities.

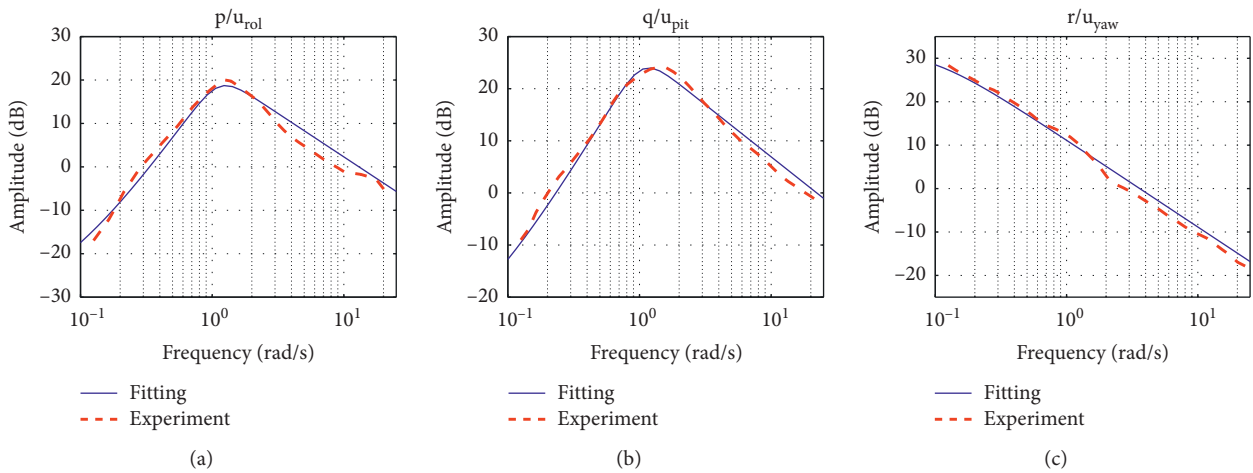


FIGURE 9: Continued.



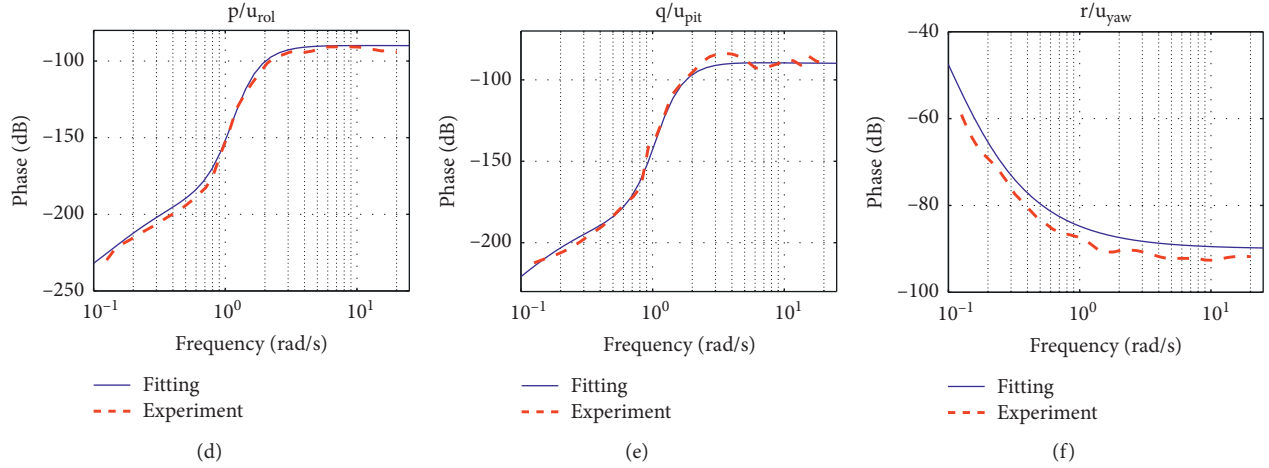


FIGURE 9: Frequency response identification results of angular rates.

$$\left. \begin{array}{l} \mathbf{A} = \begin{pmatrix} -0.0876 & 0 & 0 & 0 & 0 & 0 & 0 & -9.8010 & 0 \\ 0 & -0.0876 & 0 & 0 & 0 & 0 & 9.8010 & 0 & 0 \\ -0.1178 & -0.1172 & -1.0415 & -0.0012 & -0.0209 & -0.0509 & 0 & 0 & 0 \\ 0 & -0.6801 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0940 & 0 & 0 & 0 & -1.0699 & 0.0132 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.0122 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \\ \mathbf{B} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -20.9312 & 0 & 0.0547 & -0.0015 \\ 0 & 8.3494 & 0 & 0.0260 \\ 0 & -0.0025 & 10.0876 & -0.0281 \\ 0 & 0.0016 & 0.1337 & 3.3981 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \right\} \quad (19)$$

In order to verify the accuracy of the identified model, a series of flight tests are carried out in time domain, as shown in Figure 10. The low-pass filtered bipolar square wave signal, which is different from the identification excitation signal, is used as the excitation signal in each channel, and the evaluation of the model is realized by comparing the output of the identified model simulation and the actual test measurement. Notice that the angular rate errors in Figure 10 are large while other variables match well, because the dynamic characteristics of angular rate are ignored in high frequency in the linearized model and the angular rate

channel has relatively large measurement noise. According to [35], Theil inequality coefficient (TIC) can be used to estimate the model fitting accuracy, which is calculated as

$$J_{\text{TIC}} = \frac{\sqrt{1/n \sum_{i=1}^n (\mathbf{y}_{\text{test}} - \mathbf{y}_{\text{sim}})^T (\mathbf{y}_{\text{test}} - \mathbf{y}_{\text{sim}})}}{\sqrt{1/n \sum_{i=1}^n \mathbf{y}_{\text{sim}}^T \mathbf{y}_{\text{sim}} + 1/n \sum_{i=1}^n \mathbf{y}_{\text{test}}^T \mathbf{y}_{\text{test}}}}, \quad (20)$$

where  $\mathbf{y}_{\text{sim}}$  and  $\mathbf{y}_{\text{test}}$  are the output vector of identified model simulation and actual flight test, and  $n$  is the number of sampling points. A smaller value of TIC indicates that the accuracy of the identified model is higher, and a value less

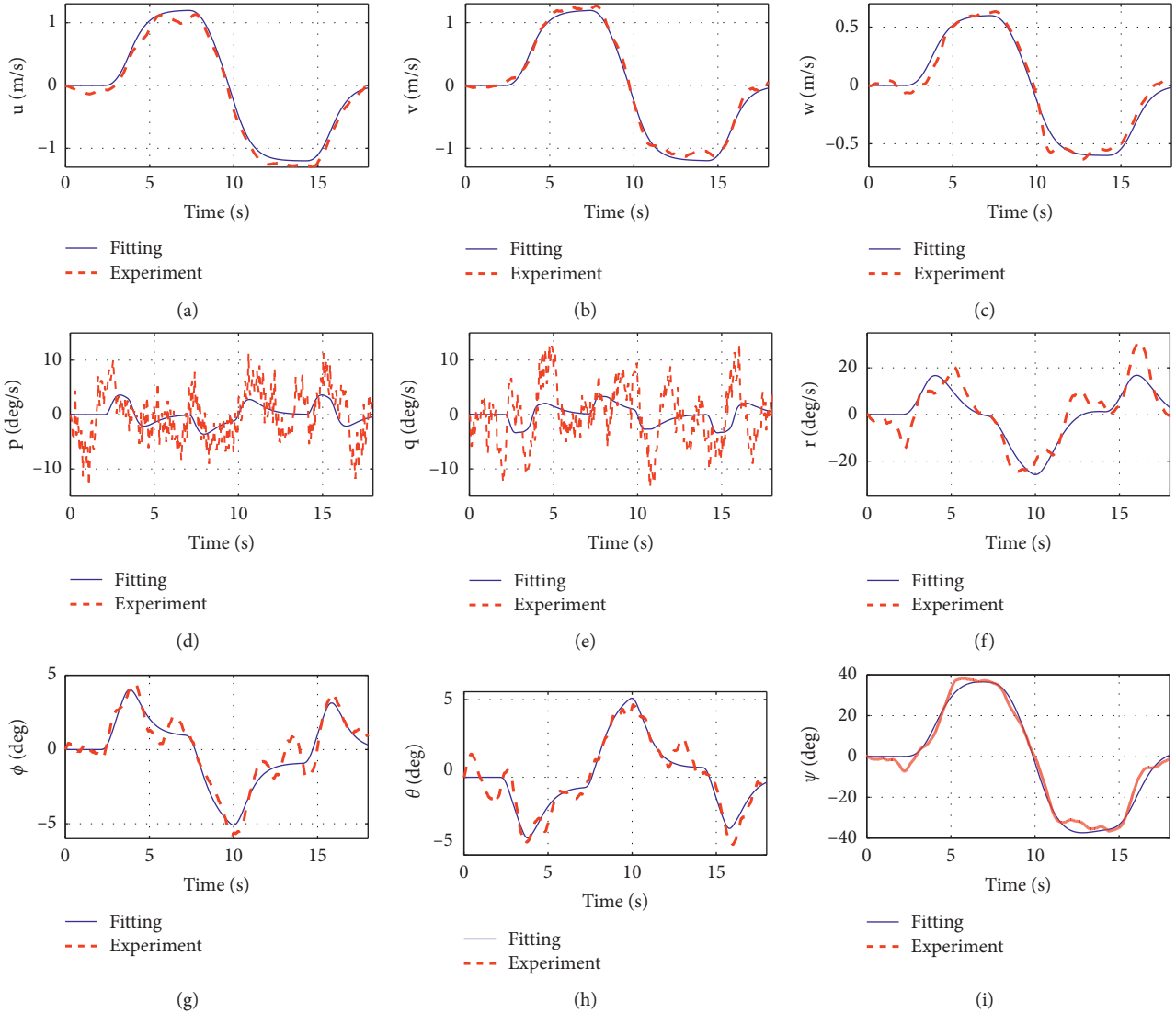


FIGURE 10: Time-domain verification of the identified model.

than 0.25 indicates that the model has good predictive performance [35].  $J_{TIC}$  of our model is 0.1596, indicating that the nominal model has satisfactory accuracy.

### 3. Composite Controller Design

The overall architecture of the aerial manipulator control system is shown in Figure 11. A typical application scenario of the aerial manipulator is that, given the reference position of the target object, the aerial vehicle flies near the target, and the manipulator moves, grabs the object, and then leaves. Since the process of arriving at the desired position of the aerial platform and completing the operation task of the manipulator is independent, and the control objectives, dynamic characteristics, and response speed of the aerial platform and the manipulator are different, the controllers of the aerial platform and the manipulator are designed separately for the purposes of reducing the controller complexity, reducing the hardware cost, and facilitating the

engineering practice. In this framework, the most important challenge is the aerial platform controller design under the motion of manipulator and external interaction disturbances of target object.

This section mainly focuses on the composite controller design of the aerial platform under the existence of manipulator disturbances. First, a two-layer basic controller is designed to ensure the stability, decoupling, and tracking performance of the platform. On the one hand, the aerial vehicle is a strongly coupled, nonlinear, multi-input multi-output (MIMO) system; thus, traditional controllers (such as PID controller) cannot guarantee good performance of the system. On the other hand, based on the premise of guaranteeing performance, the controller should have low complexity, clear structure, and high solution rate to be easy in practice. Considering these two aspects, a robust static feedback controller based on H-infinite synthesis and nonsmooth optimization is designed as the basic inner loop controller, which guarantees good robustness and stability

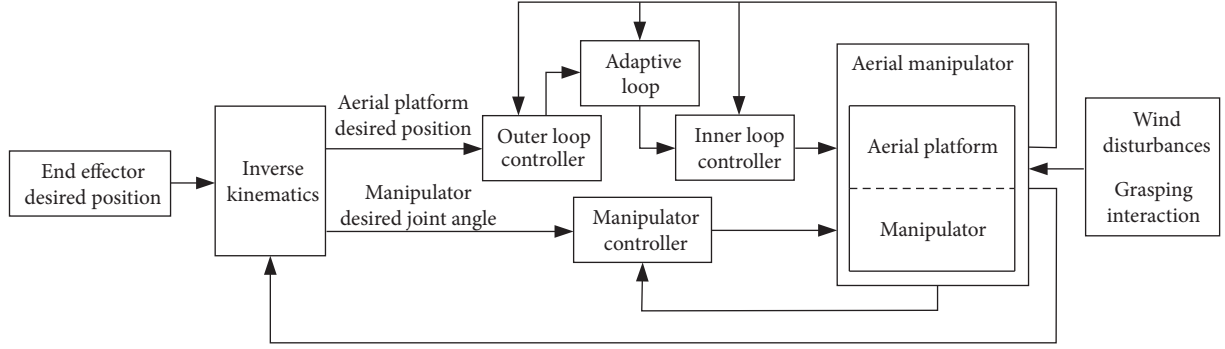


FIGURE 11: The architecture of the control system.

of the system while not increasing the system order. Second, considering the movement of the manipulator and the disturbances of the environment, an auxiliary adaptive loop is designed to estimate and compensate for the disturbances of the aerial platform. Both the static (gravity force and moment) and dynamic (inertia force and moment) disturbances of the manipulator are considered to improve the platform stability and the end effector tracking accuracy.

In addition, in order to improve the control accuracy of the manipulator under the motion of the aerial platform, a computed torque PID controller, considering the dynamics of the aerial platform, is used for manipulator control. Since the computed torque PID control is a conventional method, and the manipulator dynamics is elaborated in Section 2, the controller of the manipulator is not detailed in this paper.

**3.1. Robust Basic Controller Design.** The design objective of the basic controller is to achieve the stability and accurate position tracking of the aerial platform. The diagram of the basic controller is shown in Figure 12, which includes two loops. The inner loop is an H-infinite static full-state feedback controller  $K_b$  to ensure the decoupling and stabilization of the system, and a prepositive static controller  $K_f$  is proposed for tracking performance of the inner loop. Since the inner loop has been well decoupled and stabilized, the outer loop (position loop) is designed as a series of simple PD controllers for position tracking. According to (17), the control-oriented dynamic model of the aerial platform is

$$\begin{cases} \dot{\mathbf{x}}_b(t) = \mathbf{A}\mathbf{x}_b(t) + \mathbf{B}\mathbf{u}_b(t) + \mathbf{f}(t), \\ \mathbf{y}_b(t) = \mathbf{C}\mathbf{x}_b(t), \\ \mathbf{x}_b = [u \ v \ w \ p \ q \ r \ \varphi \ \theta \ \psi]^T, \\ \mathbf{y}_b = [u \ v \ w \ r]^T, \end{cases} \quad (21)$$

where  $\mathbf{f}$  is the disturbance vector. The reference command is  $\mathbf{r}_b = [u_{\text{ref}} \ v_{\text{ref}} \ w_{\text{ref}} \ r_{\text{ref}}]^T$ . Accordingly, the outer loop reference command is  $\mathbf{p}_r = [x_{\text{ref}} \ y_{\text{ref}} \ z_{\text{ref}} \ \psi_{\text{ref}}]^T$ . The kinematics module in Figure 12 converts the variables between the body-fixed coordinate frame and earth-fixed coordinate frame according to equation (1).

Focusing on the inner loop controller, we can define the output error as

$$\mathbf{e}_b = \mathbf{r}_b - \mathbf{y}_b. \quad (22)$$

Rewrite the system model in frequency domain and augment it, and the structured decoupling standard form of H-infinite synthesis problem [36] can be obtained as in Figure 13:

$$\begin{cases} \mathbf{y}_{\text{aug}} = [\mathbf{y}_b \ \mathbf{x}_{\text{aug}}]^T = \mathbf{G}_{\text{aug}}\mathbf{u}_{\text{aug}}, \\ \mathbf{x}_{\text{aug}} = [\mathbf{e}_b \ \mathbf{x}_b]^T, \\ \mathbf{u}_{\text{aug}} = [\mathbf{f} \ \mathbf{r}_b \ \mathbf{u}_b]^T, \\ \mathbf{G}_{\text{aug}}(\omega) = \begin{bmatrix} \mathbf{C}\mathbf{G}_f & 0 & \mathbf{C}\mathbf{G}_b \\ -\mathbf{C}\mathbf{G}_f & \mathbf{I} & -\mathbf{C}\mathbf{G}_b \\ \mathbf{G}_f & 0 & \mathbf{G}_b \end{bmatrix}, \end{cases} \quad (23)$$

where the control law of the robust controller can be calculated as

$$\begin{cases} \mathbf{u}_b = \mathbf{K}_f\mathbf{e}_b - \mathbf{K}_b\mathbf{x}_b = \mathbf{K}_{\text{st}}\mathbf{x}_{\text{aug}}, \\ \mathbf{K}_{\text{st}} = \begin{bmatrix} \mathbf{K}_f & 0 \\ 0 & -\mathbf{K}_b \end{bmatrix}. \end{cases} \quad (24)$$

For the augmented system, the closed-loop transfer function from the input  $m$  to the output  $n$  can be calculated by linear fractional transformation (LFT) as

$$\mathbf{T}_{m \rightarrow n}(s) = \mathbf{F}_l(\mathbf{G}_{\text{aug}}, \mathbf{K}_{\text{st}}). \quad (25)$$

Then, based on H-infinite synthesis method, the controller tuning process can be converted into an optimization problem as

$$\begin{aligned} \min_{\mathbf{K}_{\text{st}}} \quad & \mathbf{J} = \|\mathbf{W}_{\Delta}\mathbf{T}_{\Delta}(\mathbf{G}_{\text{aug}}, \mathbf{K}_{\text{st}})\|_{\infty} \\ \text{subject to} \quad & \mathbf{K}_{\text{st}} \text{ stabilizes } \mathbf{G}_{\text{aug}} \text{ internally} \end{aligned} \quad (26)$$

$$\|\mathbf{W}_{\Delta}\mathbf{T}_{\Delta}(\mathbf{G}_{\text{aug}}, \mathbf{K}_{\text{st}})\|_{\infty} < 1,$$

where  $\|\mathbf{W}_{\Delta}\mathbf{T}_{\Delta}\|_{\infty}$  are the performance constraint functions expressed by H-infinite norm and  $\mathbf{W}_{\Delta}$  are the corresponding weighting functions. The solvability of the optimization problem guarantees the robust stability of the closed-loop system. For the aerial platform in this paper, the following functions are carried out for ensuring tracking performance,

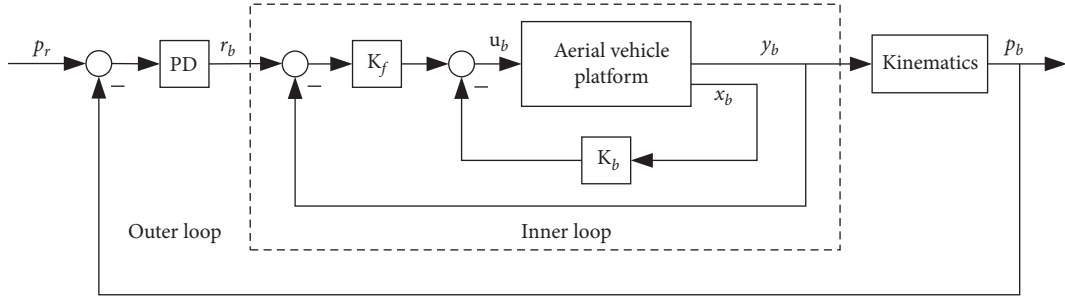


FIGURE 12: Basic controller architecture.

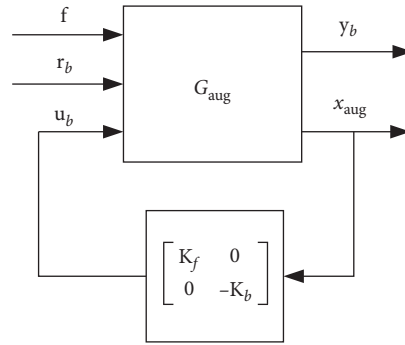


FIGURE 13: Standard structure of H-infinity synthesis.

disturbance rejection performance, and input energy limitation, respectively:

$$\begin{cases} \|W_e(s)T_{r_b} \rightarrow e_b(s)\|_{\infty} < 1, \\ \|W_b(s)T_{r_b} \rightarrow y_b(s)\|_{\infty} < 1, \\ \|W_f(s)T_{f \rightarrow y_b}(s)\|_{\infty} < 1, \\ \|W_u(s)T_{r_b} \rightarrow u_b(s)\|_{\infty} < 1. \end{cases} \quad (27)$$

Solve equation (26) using nonsmooth optimization algorithm [37]; the controller parameters are as follows. In addition, after the inner loop controller has been solved, the outer loop PD controller can be tuned easily by optimization, and the results are also given here:

$$\begin{cases} \mathbf{K}_b = \begin{bmatrix} 0.0056 & 0.0055 & 0.0498 & 0.0012 & -0.0001 & 0.0006 & 0.0086 & -0.0040 & 0.0001 \\ 0.0001 & -0.1648 & 0.0002 & 1.1658 & 0.0111 & 0.2514 & 9.3447 & -0.0118 & -0.0008 \\ 0.0751 & -0.0001 & -0.0001 & 0.0237 & 0.8012 & 0.0328 & 0.0148 & 7.3617 & -0.0003 \\ -0.0050 & 0.0078 & 0.0002 & 0.1888 & 0.0257 & 2.6449 & -0.8605 & -0.5168 & 0.0005 \end{bmatrix}, \\ \mathbf{K}_f = \begin{bmatrix} -0.0478 & 0.0009 & 0.0005 & -0.0019 \\ -0.0034 & 0.9542 & 0.0009 & 0.2473 \\ 0.0007 & 0.0049 & -0.7502 & 0.0351 \\ -0.0167 & -0.0700 & 0.0513 & 2.6541 \end{bmatrix}, \\ \text{PD} = \text{diag}(0.85 + 0.8s \quad 0.85 + 0.8s \quad 0.95 + 1.0s \quad 0.95 + 1.0s). \end{cases} \quad (28)$$

Figure 14 illustrates the robust stability margin of the closed-loop system, where the gain margin exceeds 6 dB, and the phase margin exceeds 45°. The results demonstrate that

the proposed robust controller can guarantee the basic performance of the system.

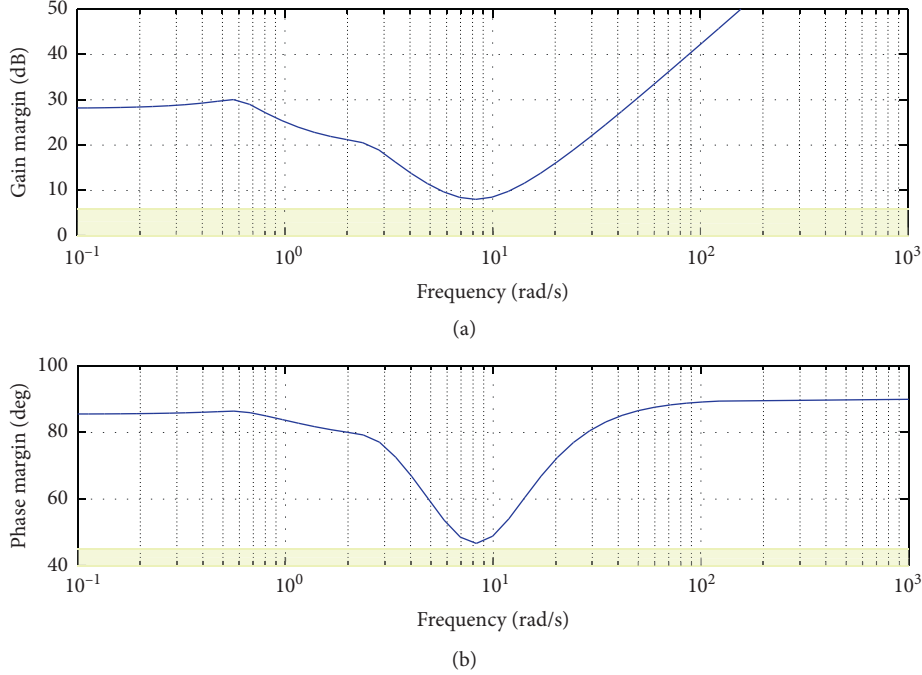


FIGURE 14: The robust stability margin with the basic controller.

**3.2. Adaptive Auxiliary Controller Design.** Although the designed robust controller has achieved basic stability and tracking performance and has antidisturbance ability to some extent, it cannot deal with the aerial manipulator system well because the disturbance caused by manipulator operation far exceeds the stability margin of the robust controller, which makes the system performance deteriorate dramatically. Therefore, it is necessary to design an adaptive auxiliary loop to compensate for the impact of the manipulator. The architecture of the adaptive controller is shown in Figure 15. The input of the adaptive loop is the output of the outer loop basic PD controller, and its output is the reference input of the inner closed-loop system. Because the inner loop robust controller is static, it does not increase or decrease the state variables and does not change the physical meanings of the system states; thus, the inner closed-loop system can be obtained easily from Section 3.1 as

$$\begin{cases} \dot{\mathbf{x}}_b(t) = \mathbf{A}_{\text{ad}}\mathbf{x}_b(t) + \mathbf{B}_{\text{ad}}\mathbf{u}_{\text{ad}}(t) + \mathbf{f}(t), \\ \mathbf{y}_b(t) = \mathbf{C}\mathbf{x}_b(t), \end{cases} \quad (29)$$

which is clearly bounded-input bounded-state (BIBS) stable relying on the condition of (26) in the inner loop controller design process, and  $\mathbf{A}_{\text{ad}}$  is Hurwitz matrix. Besides, according to the previous description, all the state variables are measurable and  $\mathbf{x}_b(0) = \mathbf{0}$ .  $\mathbf{f}(t)$  is the time-varying disturbance, which is bounded and differentiable, and also the differential is bounded. The boundary of  $\mathbf{f}(t)$  can be estimated by the manipulator dynamics and max grasping payload introduced in Section 2, with some redundancy on this basis. Thus, the following condition is given:

$$\begin{cases} \exists \text{compact set } \mathbb{F}, & \forall t \geq 0, \mathbf{f}(t) \in \mathbb{F}, \\ \exists \mathcal{F} \text{ and } d_F, & \forall t \geq 0, \|\mathbf{f}(t)\|_2 \leq \mathcal{F}, \|\dot{\mathbf{f}}(t)\|_2 \leq d_F, \end{cases} \quad (30)$$

The adaptive loop consists of three parts: first, a state predictor is designed for state prediction; then, the adaptive law is designed for disturbance estimation based on the prediction error; finally, the control law is reconstructed to compensate for the disturbance and track the desired reference. In particular, inspired by the theory in [38], a filter matrix  $\mathbf{C}(s)$  is applied to shape the control input, which decouples the adaptive rate from system robustness. Thus, the large adaptive gain can be adopted while avoiding adverse effects on the system states, which ensures the accurate estimation of disturbance and guarantees good transient and steady-state performance of the adaptive controller.

**3.2.1. Design of the Adaptive Auxiliary Controller.** Consider the following state predictor of the system in equation (29):

$$\begin{cases} \dot{\hat{\mathbf{x}}}_b(t) = \mathbf{A}_{\text{ad}}\hat{\mathbf{x}}_b(t) + \mathbf{B}_{\text{ad}}\mathbf{u}_{\text{ad}}(t) + \hat{\mathbf{f}}(t), \\ \hat{\mathbf{y}}_b(t) = \mathbf{C}\hat{\mathbf{x}}_b(t), \end{cases} \quad (31)$$

where the cap  $\hat{\cdot}$  denotes the estimation of the corresponding variable. Then, the estimation error dynamics can be calculated from (29) and (31) as

$$\begin{cases} \dot{\tilde{\mathbf{x}}}_b(t) = \mathbf{A}_{\text{ad}}\tilde{\mathbf{x}}_b(t) + \tilde{\mathbf{f}}(t), \\ \tilde{\mathbf{x}}_b(t) = \hat{\mathbf{x}}_b(t) - \mathbf{x}_b(t), \\ \tilde{\mathbf{f}}(t) = \hat{\mathbf{f}}(t) - \mathbf{f}(t). \end{cases} \quad (32)$$

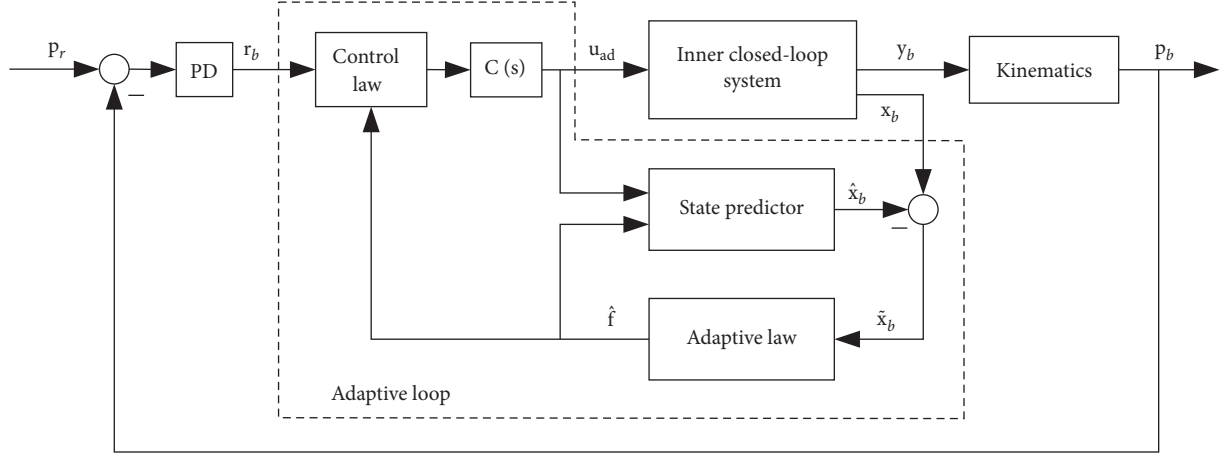


FIGURE 15: Adaptive controller architecture.

In order to ensure the stability of estimation error, consider the following candidate Lyapunov function:

$$V(\tilde{x}_b(t), \tilde{f}(t)) = \tilde{x}_b^T(t) \mathbf{P} \tilde{x}_b(t) + \Gamma^{-1} \tilde{f}^T(t) \tilde{f}(t), \quad (33)$$

where  $\Gamma$  is the adaptive gain. Since  $A_{ad}$  is Hurwitz,  $\mathbf{P}$  is the solution of the following algebraic Lyapunov equation:

$$\mathbf{A}_{ad}^T \mathbf{P} + \mathbf{P} \mathbf{A}_{ad} = -\mathbf{I}. \quad (34)$$

The derivation of (33) is

$$\begin{aligned} \dot{V}(t) &= \dot{\tilde{x}}_b^T(t) \mathbf{P} \tilde{x}_b(t) + \tilde{x}_b^T(t) \mathbf{P} \dot{\tilde{x}}_b(t) + \Gamma^{-1} \left( \dot{\tilde{f}}^T(t) \tilde{f}(t) + \tilde{f}^T(t) \dot{\tilde{f}}(t) \right) \\ &= -\tilde{x}_b^T(t) \tilde{x}_b(t) + 2\tilde{f}^T(t) \left( (\tilde{x}_b^T(t) \mathbf{P})^T + \Gamma^{-1} \dot{\tilde{f}}(t) \right). \end{aligned} \quad (35)$$

Based on Lyapunov stability criterion, let the adaptive law be

$$\dot{\tilde{f}}(t) = \Gamma \text{Proj} \left( \tilde{f}(t), -(\tilde{x}_b^T(t) \mathbf{P})^T \right), \quad (36)$$

where Proj denotes the projection operator, which prevents the divergence and ensures the boundedness of the adaptive parameter, detailed in [39]. Thus, (35) can be further derived as

$$\dot{V}(t) = -\tilde{x}_b^T(t) \tilde{x}_b(t) \leq 0. \quad (37)$$

Notice that in the adaptive law derivation in (35), it is assumed that the change rate of the disturbance  $f$  is slower than the convergence rate of the adaptive estimation; thus, its derivative is approximately 0. In the performance analysis of the next section, this assumption will be relaxed and more general situations will be considered. The estimation error stability under arbitrary disturbance including arbitrary fast time-varying disturbance will be proved in detail with the proposed adaptive law in (36).

Then the control law in Laplace form can be calculated as follows (refer to Figure 15):

$$\mathbf{u}_{ad}(s) = \mathbf{C}(s) \left( \mathbf{K}_g \mathbf{r}_b(s) - \mathbf{F}_1(s) \hat{f}(s) \right). \quad (38)$$

where  $\mathbf{K}_g = -(\mathbf{C} \mathbf{A}_{ad}^{-1} \mathbf{B}_{ad})^{-1}$ ,  $\mathbf{F}_1(s) = \mathbf{H}_1^{-1}(s) \mathbf{C} \mathbf{H}_0(s)$ , and  $\mathbf{H}_0(s) = (s\mathbf{I} - \mathbf{A}_{ad})^{-1}$ ,  $\mathbf{H}(s) = \mathbf{H}_0(s) \mathbf{B}_{ad}$ ,  $\mathbf{H}_1(s) = \mathbf{C} \mathbf{H}(s)$ .  $\mathbf{C}(s)$  is strictly proper and stable low-pass filter matrix. Following the idea in [38],  $\mathbf{C}(s)$  could decouple the adaptive rate from system robustness, and the selection of it should satisfy the following conditions:

$$\begin{cases} \mathbf{C}(s) \text{ is strictly proper and stable with } \mathbf{C}(0) = \mathbf{I}, \\ \exists \rho_{r1}, \|\mathbf{C}(s) \mathbf{H}_1^{-1}(s)\|_{L_1} < \rho_{r1}, \\ \exists \rho_{r2}, \|\mathbf{H}_0(s) (\mathbf{I} - \mathbf{B}_{ad} \mathbf{C}(s) \mathbf{F}_1(s))\|_{L_1} < \rho_{r2}, \end{cases} \quad (39)$$

where  $\|\Delta(s)\|_{L_1}$  refers to the  $L_1$  norm of  $\Delta(s)$  and the last two conditions indicate that the corresponding  $L_1$  norm gain is finite.

Thus, the design process of the adaptive controller is completed. To sum up, for the system in (29), choose the appropriate adaptive gain  $\Gamma$  and filter matrix  $\mathbf{C}(s)$  with satisfying condition (39); the control input is calculated via (31), (36), and (38) as

$$\begin{cases} \dot{\hat{x}}_b(t) = \mathbf{A}_{ad} \hat{x}_b(t) + \mathbf{B}_{ad} \mathbf{u}_{ad}(t) + \hat{f}(t), \\ \dot{\hat{f}}(t) = \Gamma \text{Proj} \left( \hat{f}(t), -(\hat{x}_b^T(t) \mathbf{P})^T \right), \\ \mathbf{u}_{ad}(s) = \mathbf{C}(s) \left( \mathbf{K}_g \mathbf{r}_b(s) - \mathbf{F}_1(s) \hat{f}(s) \right). \end{cases} \quad (40)$$

**3.2.2. Performance Analysis of the Adaptive Auxiliary Controller.** Consider the ideal form of the adaptive controller, which is assumed to compensate for the disturbance perfectly. The closed-loop reference system is defined as

$$\begin{cases} \dot{\mathbf{x}}_{ref}(t) = \mathbf{A}_{ad} \mathbf{x}_{ref}(t) + \mathbf{B}_{ad} \mathbf{u}_{ref}(t) + \mathbf{f}(t), \\ \mathbf{y}_{ref}(t) = \mathbf{C} \mathbf{x}_{ref}(t), \\ \mathbf{u}_{ref}(s) = \mathbf{C}(s) \left( \mathbf{K}_g \mathbf{r}_b(s) - \mathbf{F}_1(s) \mathbf{f}(s) \right), \end{cases} \quad (41)$$

and  $\mathbf{x}_{\text{ref}}(0) = 0$ . Using Laplace transformation, the closed-loop system (41) can be described as

$$\mathbf{x}_{\text{ref}}(s) = \mathbf{H}(s)\mathbf{C}(s)\mathbf{K}_g\mathbf{r}_b(s) + \mathbf{H}_0(s)(\mathbf{I} - \mathbf{B}_{\text{ad}}\mathbf{C}(s)\mathbf{F}_1(s))\mathbf{f}(s). \quad (42)$$

The following lemma describes the L1 norm stability theorem [38].

**Lemma 1.** *A continuous-time linear time-invariant (LTI) MIMO system  $y(s) = G(s)u(s)$  is BIBO stable if its L1 norm is bounded and for arbitrary bounded  $u(t)$  one has*

$$\|\mathbf{y}\|_{L_\infty} \leq \|\mathbf{G}(s)\|_{L_1} \|\mathbf{u}\|_{L_\infty}, \quad (43)$$

where  $\|\Delta(s)\|_{L_1}$  refers to the  $L_1$  norm of  $\Delta(s)$ , and  $\|\Delta(s)\|_{L_\infty}$  refers to the  $L_\infty$  norm of  $\Delta(s)$ .

$$\begin{aligned} \dot{V}(t) &= \dot{\tilde{\mathbf{x}}}_b^T(t)\mathbf{P}\tilde{\mathbf{x}}_b(t) + \tilde{\mathbf{x}}_b^T(t)\mathbf{P}\dot{\tilde{\mathbf{x}}}_b(t) + \Gamma^{-1}\left(\dot{\tilde{\mathbf{f}}}^T(t)\tilde{\mathbf{f}}(t) + \tilde{\mathbf{f}}^T(t)\dot{\tilde{\mathbf{f}}}(t)\right) \\ &= -\tilde{\mathbf{x}}_b^T(t)\tilde{\mathbf{x}}_b(t) + 2\tilde{\mathbf{f}}^T(t)\left(\left(\tilde{\mathbf{x}}_b^T(t)\mathbf{P}\right)^T + \Gamma^{-1}\dot{\tilde{\mathbf{f}}}(t)\right) - 2\Gamma^{-1}\tilde{\mathbf{f}}^T(t)\dot{\tilde{\mathbf{f}}}(t) \\ &\leq -\tilde{\mathbf{x}}_b^T(t)\tilde{\mathbf{x}}_b(t) + 2\tilde{\mathbf{f}}^T(t)\left(\left(\tilde{\mathbf{x}}_b^T(t)\mathbf{P}\right)^T + \mathbf{Proj}\left(\tilde{\mathbf{f}}(t), -\left(\tilde{\mathbf{x}}_b^T(t)\mathbf{P}\right)^T\right)\right) + 2\Gamma^{-1}\left|\tilde{\mathbf{f}}^T(t)\dot{\tilde{\mathbf{f}}}(t)\right| \\ &\leq -\tilde{\mathbf{x}}_b^T(t)\tilde{\mathbf{x}}_b(t) + 4\Gamma^{-1}d_{\mathcal{F}}\mathcal{F}. \end{aligned} \quad (45)$$

Also, from (30) and (33) one has

$$\begin{cases} \max_{t \geq 0} \left(\tilde{\mathbf{f}}^T(t)\tilde{\mathbf{f}}(t)\right) \leq 4 \max_{\mathbf{f} \in \mathcal{F}} \|\mathbf{f}\|_2^2 = 4F^2, \\ V(0) = \Gamma^{-1}\tilde{\mathbf{f}}^T(0)\tilde{\mathbf{f}}(0) \leq 4\Gamma^{-1}F^2. \end{cases} \quad (46)$$

Let  $\Omega_m := 4\lambda_{\max}(\mathbf{P})d_{\mathcal{F}}F + 4F^2$ ; if there exist some  $t > 0$  such that  $V(t) > \Gamma^{-1}\Omega_m$ , then, it follows from (33) and (46) that

$$\tilde{\mathbf{x}}_b^T(t)\mathbf{P}\tilde{\mathbf{x}}_b(t) > 4\Gamma^{-1}\lambda_{\max}(\mathbf{P})d_{\mathcal{F}}\mathcal{F}, \quad (47)$$

and thus,

$$\tilde{\mathbf{x}}_b^T(t)\tilde{\mathbf{x}}_b(t) \geq \lambda_{\min}^{-1}(\mathbf{P})\left(\tilde{\mathbf{x}}_b^T(t)\mathbf{P}\tilde{\mathbf{x}}_b(t)\right) > 4\Gamma^{-1}d_{\mathcal{F}}\mathcal{F}. \quad (48)$$

Then, from equations (45) and (48), one can deduce that

$$\dot{V}(t) < 0. \quad (49)$$

According to equation (46), it is derived that  $V(0) < \Gamma^{-1}\Omega_m$ . Based on the deduction above, it can be obtained that

$$\forall t \geq 0, \quad V(t) \leq \Gamma^{-1}\Omega_m. \quad (50)$$

Then, the bound of the state estimation error can be derived as

$$\begin{aligned} \lambda_{\min}(\mathbf{P})\|\tilde{\mathbf{x}}_b(t)\|_2^2 &\leq \tilde{\mathbf{x}}_b^T(t)\mathbf{P}\tilde{\mathbf{x}}_b(t) \leq V(t) \leq \Gamma^{-1}\Omega_m \\ &\Rightarrow \|\tilde{\mathbf{x}}_b(t)\|_2 \leq \sqrt{\lambda_{\min}^{-1}(\mathbf{P})\Gamma^{-1}\Omega_m}, \end{aligned} \quad (51)$$

Since  $A_{\text{ad}}$  is Hurwitz and the conditions in (30) and (39) are satisfied, for all  $t \geq 0$ , the L1 norm of system (42) is finite; thus,  $\|\mathbf{x}_{\text{ref}}\|_{L_\infty}$  is uniformly bounded, the closed-loop reference system (41) is BIBS stable, and the reference system has the following performance bound:

$$\|\mathbf{x}_{\text{ref}}\|_{L_\infty} \leq \|\mathbf{H}(s)\mathbf{C}(s)\mathbf{K}_g\|_{L_1}\|\mathbf{r}_b\|_{L_\infty} + \|\mathbf{H}_0(s)(\mathbf{I} - \mathbf{B}_{\text{ad}}\mathbf{C}(s)\mathbf{F}_1(s))\|_{L_1}\|\mathbf{f}\|_{L_\infty}. \quad (44)$$

For the defined state predictor (31), consider the candidate Lyapunov function (33) again; its derivation under arbitrary fast time-varying disturbance with the adaptive law (40) can be derived as

which proves that the estimation error is uniformly bounded and the state predictor is Lyapunov stable. It also can be drawn from equation (51) that the performance of the state predictor is inversely proportional to the square root of the adaptive gain  $\Gamma$ , which can be improved by increasing the adaptive gain.

It follows from equations (29) and (40) that

$$\begin{cases} \mathbf{x}_b(s) = \mathbf{H}(s)\mathbf{C}(s)\mathbf{K}_g\mathbf{r}_b(s) + \mathbf{H}_0(s) - (\mathbf{I} - \mathbf{B}_{\text{ad}}\mathbf{C}(s)\mathbf{F}_1(s)) \\ \mathbf{f}(s) - \mathbf{H}(s)\mathbf{C}(s)\mathbf{F}_1(s)\tilde{\mathbf{f}}(s), \\ \mathbf{u}_{\text{ad}}(s) = \mathbf{C}(s)\left(\mathbf{K}_g\mathbf{r}_b(s) - \mathbf{F}_1(s)\mathbf{f}(s) - \mathbf{F}_1(s)\tilde{\mathbf{f}}(s)\right). \end{cases} \quad (52)$$

Consider equations (32), (42), and (52) simultaneously; the adaptive state error can be calculated as

$$\mathbf{x}_{\text{ref}}(s) - \mathbf{x}_b(s) = \mathbf{H}(s)\mathbf{C}(s)\mathbf{H}_1^{-1}(s)\mathbf{C}\tilde{\mathbf{x}}_b(s). \quad (53)$$

According to the condition in equation (39) and the bound in equation (51), for all  $t \geq 0$ , the adaptive state error is uniformly bounded:

$$\begin{aligned} \|\mathbf{x}_{\text{ref}} - \mathbf{x}_b\|_{L_\infty} &\leq \|\mathbf{H}(s)\mathbf{C}(s)\mathbf{H}_1^{-1}(s)\mathbf{C}\|_{L_1}\|\tilde{\mathbf{x}}_b\|_{L_\infty} \\ &\leq \|\mathbf{H}(s)\mathbf{C}(s)\mathbf{H}_1^{-1}(s)\mathbf{C}\|_{L_1}\sqrt{\lambda_{\min}^{-1}(\mathbf{P})\Gamma^{-1}\Omega_m}. \end{aligned} \quad (54)$$

The adaptive input error can be calculated by equations (41) and (52), and in a similar way to the state error, the bound of adaptive input error for all  $t \geq 0$  can be derived as

$$\begin{aligned} \|\mathbf{u}_{\text{ref}} - \mathbf{u}_{\text{ad}}\|_{L_\infty} &\leq \|C(s)\mathbf{H}_1^{-1}(s)C\|_{L_1} \|\tilde{x}_b\|_{L_\infty} \\ &\leq \|C(s)\mathbf{H}_1^{-1}(s)C\|_{L_1} \sqrt{\lambda_{\min}^{-1}(\mathbf{P})\Gamma^{-1}\Omega_m}. \end{aligned} \quad (55)$$

Since the closed-loop reference system proved to be BIBS stable in (44), combined with (54) and (55), it can be derived that the closed-loop system with the proposed adaptive controller is BIBS stable, and the performance bound relies on the adaptive gain  $\Gamma$ .

Because the system robustness is decoupled with the adaptive gain by the filter matrix  $C(s)$ , a large  $\Gamma$  can be applied to guarantee good transient and steady-state performance. Theoretically, the choice of adaptive gain is limited only by the performance of the on-board processor hardware. In some literature works, the disturbance  $f$  is called unmatched uncertainty, and, for a general open-loop MIMO system, only its effect on the system output (not the system state) can be compensated for by the control input. However, our system for adaptive controller design is BIBS stable and the system states are equal to system outputs ( $C$  is an identity matrix), which are measurable with actual physical meanings, and the disturbance boundary can be obtained by the manipulator dynamics. Moreover, thanks to the filter matrix  $C(s)$  and the large adaptive gain, the good performance of the adaptive system can be achieved while not leading to adverse effect on the system states. Therefore, the disturbance  $f$  of our system can be estimated and compensated for well. Based on the above discussion, the parameters of the adaptive controller are given as

$$\begin{cases} C(s) = \left( \frac{25}{s^2 + 6s + 25} \right) \mathbf{I}, \\ \Gamma = 1000. \end{cases} \quad (56)$$

#### 4. Simulation Verification and Analysis

The aim of the simulation tests is to verify that the proposed aerial manipulator with the composite controller can meet the requirements of high precision operation tasks well. Concretely, they include three aspects: first, verifying the feasibility of the proposed innovative aerial manipulator design with tandem ducted fans; second, verifying the effectiveness of the proposed basic controller for the basic performance; third, comparing the stability performance and tracking performance of the aerial manipulator with and without the adaptive auxiliary loop to verify the effectiveness of the adaptive loop design.

There are three scenarios considered: (1) the basic test of the aerial platform without the manipulator moving to verify the basic performance; (2) the tracking test of the aerial platform under the random moving of the manipulator to verify the effectiveness of the adaptive auxiliary controller; (3) the three-dimensional space tracking test of the manipulator end effector with given ideal reference trajectory to verify the positioning and tracking accuracy of the gripper.

All simulation parameters follow the real physical system described in Section 2.1.

*Scenario 1.* Let the manipulator be in its initial posture, where the manipulator is vertically below the aerial vehicle (the corresponding initial joint angles are  $0^\circ$ ,  $90^\circ$ , and  $0^\circ$ , respectively). At this condition, the gravity center of the manipulator is located on the  $Z_b$ -axis of the vehicle body-fixed frame, and the manipulator is symmetrical with respect to the  $X_b$ - and  $Y_b$ -axis. Give the step references of the aerial platform position  $x$ ,  $y$ ,  $z$  and yaw angle  $\psi$  when the manipulator remains in its initial posture, and the system responses are shown in Figure 16, where BC refers to the basic controller and AC refers to the adaptive controller with the adaptive auxiliary loop.

The results show that when the manipulator is not moving, the basic controller can track the desired reference well with no overshoot and steady-state error. The response times of the longitudinal channel and the lateral channel are about 4 s, the altitude channel is about 3 s, and the yaw channel is the shortest, which is 2.5 s. The results illustrate that the innovative structure design of the aerial manipulator with tandem ducted fans is feasible, and the basic controller can achieve good performance in basic condition. The result with the adaptive auxiliary loop is also given in Figure 16, which has faster response time than the basic controller, especially in the longitudinal channel and the lateral channel, which are 2.5 s and 2 s, respectively. This is because the existence of the manipulator will produce gravity moment in the opposite direction when the vehicles roll or pitch to generate the control moment, which makes the actual control moment less than the ideal value, and the adaptive auxiliary controller can compensate for it so as to improve the response speed of the system. Since the initial posture of the manipulator coincides with the  $Z_b$ -axis of the vehicle, it has little effect on the altitude channel and yaw channel, and thus the response time of the adaptive controller is close to that of the basic controller.

*Scenario 2.* Let the manipulator joints move randomly; the tracking responses of the aerial platform position and yaw angle under the manipulator disturbances are shown in Figure 17. Similarly, BC refers to the basic controller and AC refers to the adaptive controller. The motions of the three joints of the manipulator are shown in Figure 18. Notice that the initial angles of the three joints are  $0^\circ$ ,  $90^\circ$ , and  $0^\circ$ , respectively, and the maximum ranges of the motion follow the limits in Table 2.

The results show that the system responses with the basic controller deviate significantly from the expected references and cannot meet the performance requirements under the random motion disturbances of the manipulator. The lateral channel has the largest fluctuation with a maximum close to 100%, the longitudinal channel takes the second place, and the influences of the altitude channel and the yaw channel are relatively small. This is consistent with the structural characteristics of the aerial manipulator. As shown in Figure 3, the manipulator extends from the lateral side of the aerial platform, so it mainly affects the lateral channel of the platform, and the main challenge of the control system is also in the lateral channel. In addition, when joint 1 of the manipulator rotates away from its initial angle, it will also



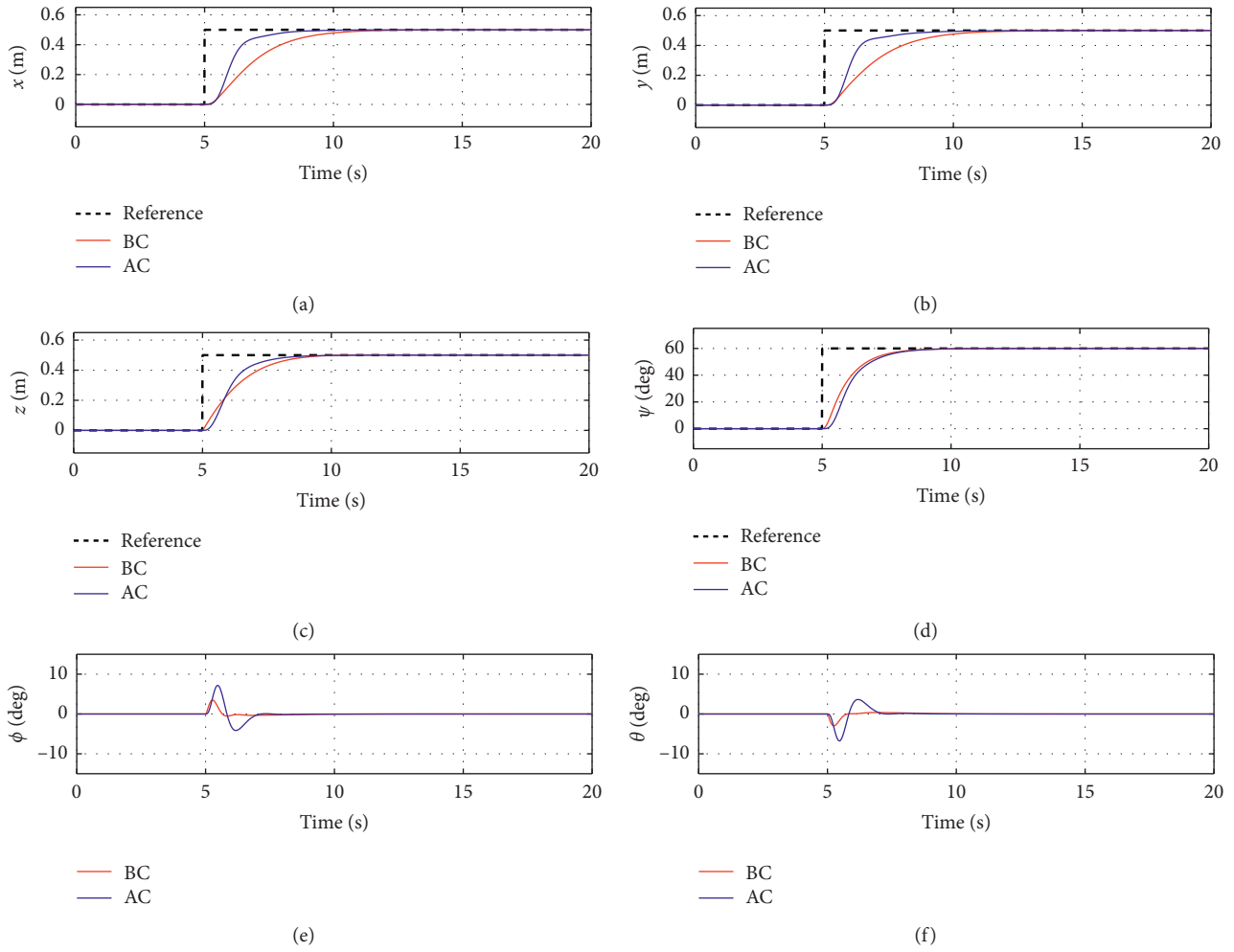


FIGURE 16: System responses in Scenario 1.

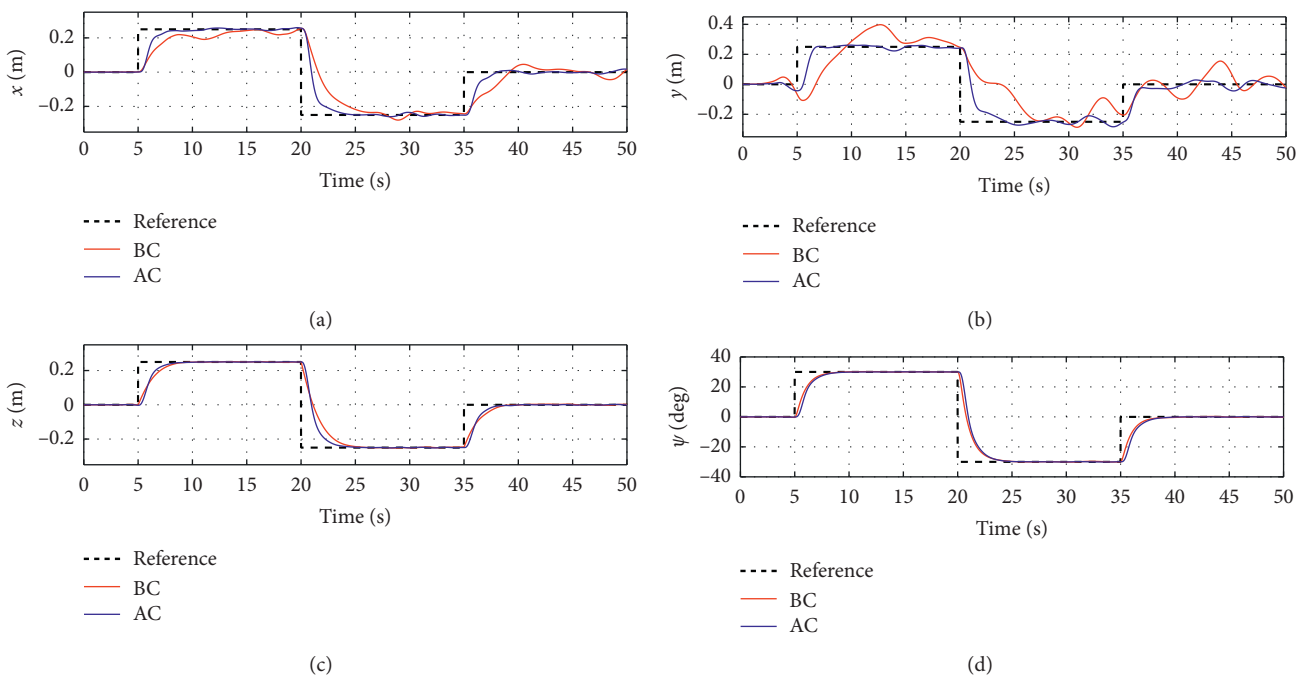


FIGURE 17: Continued.

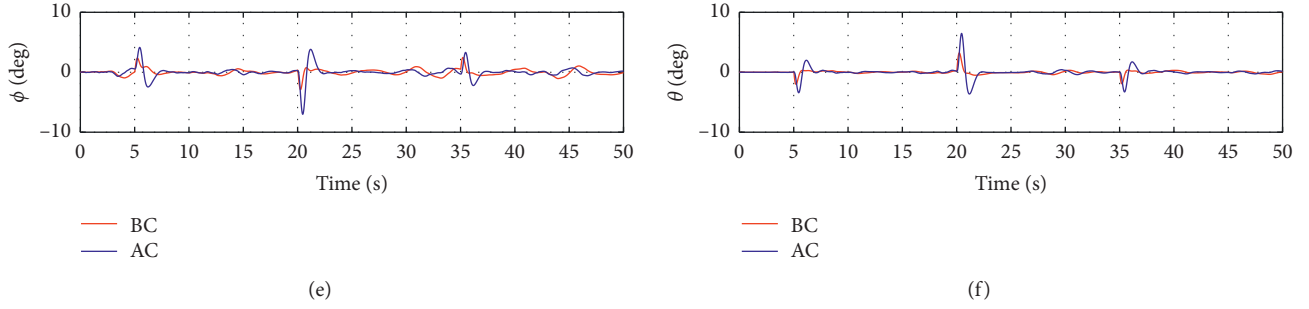


FIGURE 17: System responses in Scenario 2.

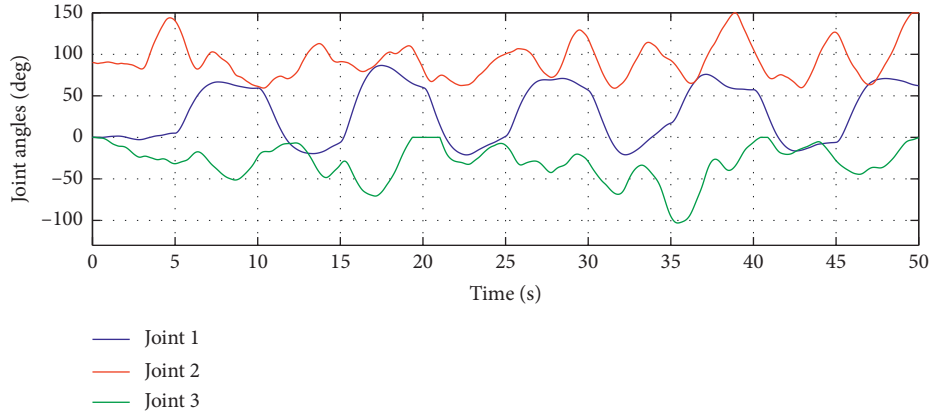


FIGURE 18: Random motions of the manipulator joints in Scenario 2.

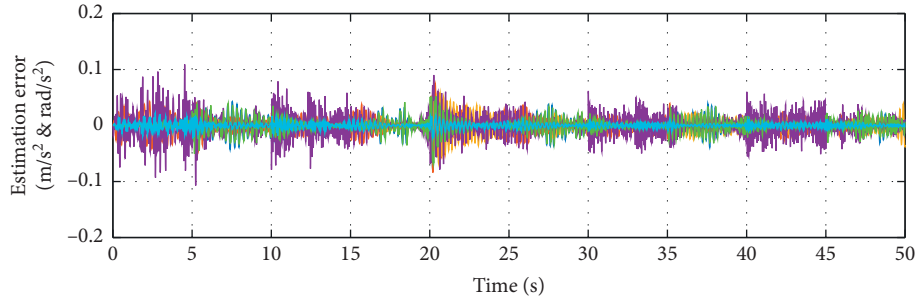


FIGURE 19: Disturbance estimation performance in Scenario 2.

TABLE 3: RMSE in Scenario 2.

| State variables     | $x$ (m) | $y$ (m) | $z$ (m) | $\psi$ (deg) |
|---------------------|---------|---------|---------|--------------|
| Basic controller    | 0.0981  | 0.1359  | 0.0759  | 8.8462       |
| Adaptive controller | 0.0782  | 0.0810  | 0.0750  | 9.0389       |

have some influence on the longitudinal channel. The adaptive controller can effectively estimate and compensate for the motion disturbance of the manipulator to ensure good disturbance rejection and tracking performance. Figure 19 also shows the disturbance estimation error of the adaptive controller, indicating that the controller has good estimation accuracy. In order to evaluate the performance of the adaptive controller more accurately, the root mean

square error (RMSE) values of the system responses under the two controllers are given in Table 3. The results show that, in the lateral channel, which has the largest fluctuation, the proposed adaptive auxiliary loop reduces the tracking error by more than 40%. The two controllers have similar performance in the altitude and yaw channel due to the correspondingly small disturbance caused by the manipulator.

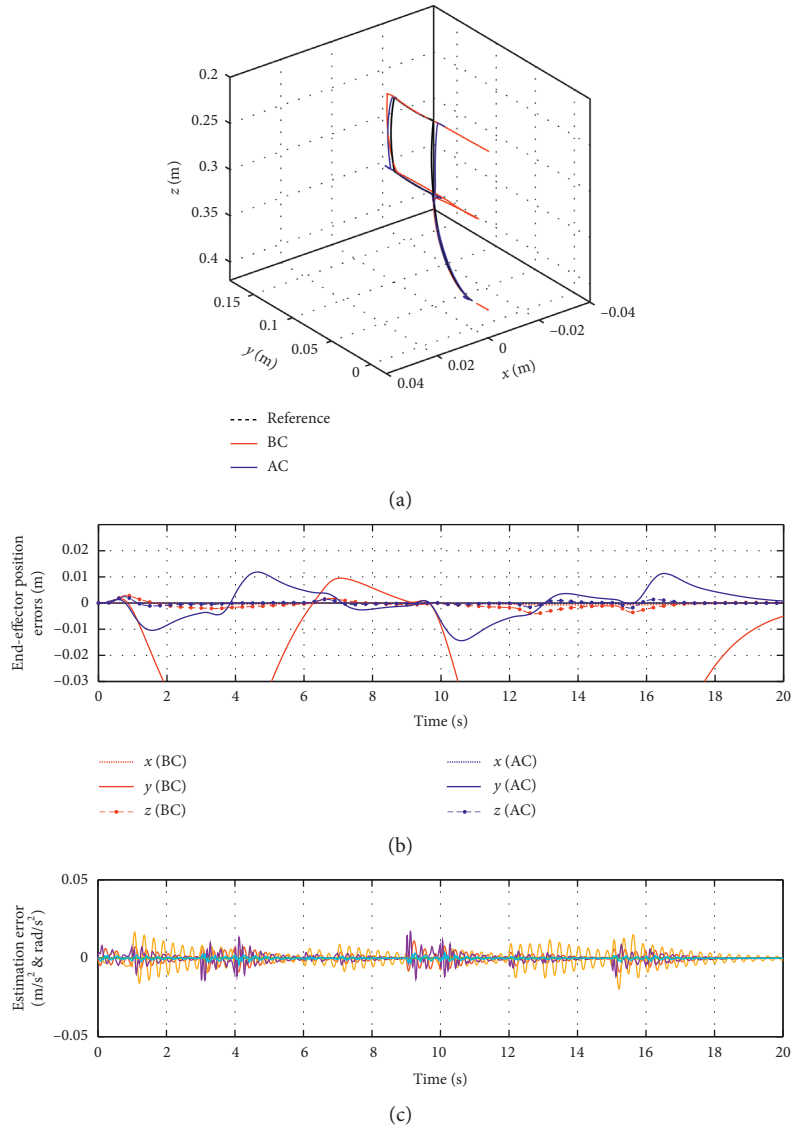


FIGURE 20: System responses in Scenario 3.1.

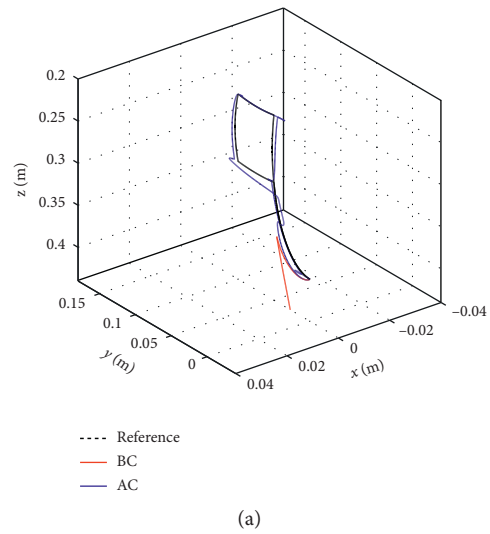


FIGURE 21: Continued.

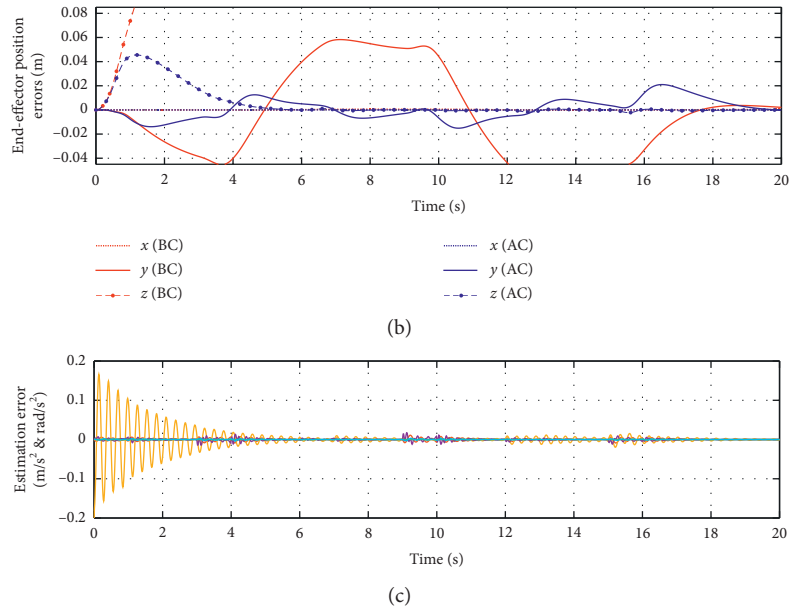


FIGURE 21: System responses in Scenario 3.2.

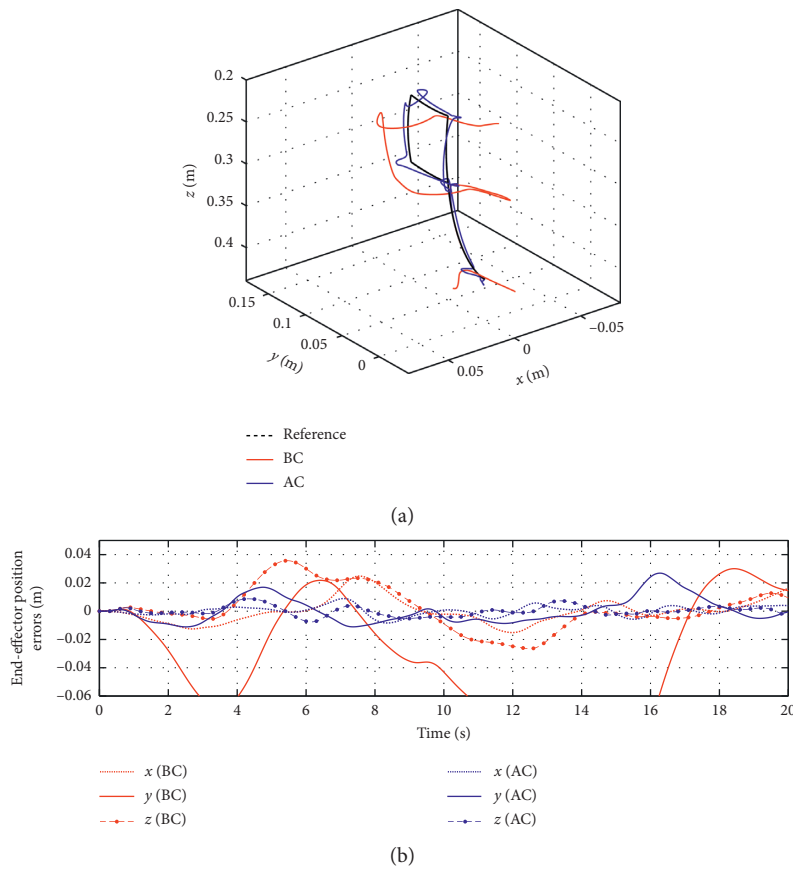


FIGURE 22: Continued.

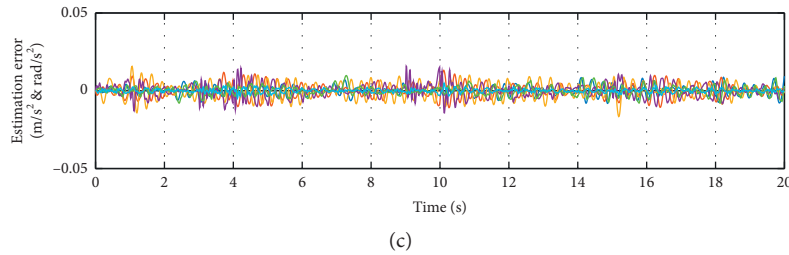


FIGURE 22: System responses in Scenario 3.3.

*Scenario 3.* Given the reference trajectory of the end effector of the manipulator in three-dimensional space, the responses of the end effector with and without the adaptive loop are shown in Figures 20–22. The reference trajectory starts from the manipulator initial posture, then passes through four reference points in the  $Y$ - $Z$  plane, with each point staying for 3 seconds, and finally returns to the initial posture. In order to fully evaluate the performance of the controller, three conditions are considered, of which Scenario 3.1 is the nominal condition, Scenario 3.2 assumes that the end effector grabs a 100 g object, and Scenario 3.3 considers gust disturbances, simulated by the Dryden model [40].

In Scenario 3.1, the adaptive controller can achieve accurate trajectory tracking of the end effector with a tracking error of approximately  $\pm 0.01$  m. The response with the basic controller has a serious deviation in the  $Y$  direction, and the system cannot completely track the desired trajectory. Since the manipulator moves in the  $Y$ - $Z$  plane, the fluctuations of the system in the  $X$  and  $Z$  directions are relatively small. In Scenario 3.2, the adaptive controller can also track the desired position, but, due to the influence of the grasping object, the overall tracking accuracy of the system decreases compared with Scenario 3.1, and its maximum fluctuation in the  $Y$  direction is about 0.02 m. A 0.04 m offset occurs in the altitude direction under the action of the object, and then the adaptive controller quickly realizes the estimation convergence and compensation. Similar to Scenario 3.1, the trajectory with the basic controller also has a serious deviation in the  $Y$  direction. Moreover, it diverges rapidly in the  $Z$  direction. In Scenario 3.3, the responses of the system in all three directions have greater fluctuations due to the influence of wind gusts. The adaptive controller can still basically achieve the tracking of the reference, and by contrast the responses with the basic controller experience severe fluctuations, which cannot meet the system requirements. In conclusion, the proposed adaptive auxiliary controller can realize precise positioning and trajectory tracking of the end effector of the aerial manipulator.

## 5. Conclusion

This paper proposes an innovative aerial manipulator based on tandem ducted fans, which can easily realize the omnidirectional manipulation of side-on and below and is particularly suitable for the tasks in confined environment. The comprehensive nonlinear dynamic model of the aerial

manipulator is established, based on which a control-oriented state space model of the aerial platform is derived using frequency domain identification method. The identification model is verified in both time and frequency domain. A composite controller of the aerial platform is designed under the existence of manipulator. First, a robust  $H$ -infinite controller is designed to ensure the basic stability, decoupling, and robust performance of the platform. Then, an auxiliary adaptive controller is designed to estimate and compensate for the effect of the manipulator dynamics. The computer simulation tests verify the feasibility of the proposed innovative aerial manipulator design and the effectiveness of the proposed controller. The results show that, with the proposed composite adaptive controller, the aerial vehicle can achieve good performance under the manipulator motion, and the manipulator end effector can realize precise positioning and trajectory tracking with the grasping load and wind gusts.

In future work, the indoor positioning system will be studied and applied, so as to lay the foundation for actual flight tests. In addition, the trajectory planning of the aerial manipulator will also be studied. Then the actual grasping test will be carried out. Furthermore, the vision-based target recognition and the interaction dynamics would be considered to achieve more complex operation tasks.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant no. 51505031) and China Scholarship Council (Grant no. 201806030063).

## References

- [1] P. Ridao, M. Carreras, D. Ribas, P. J. Sanz, and G. Oliver, "Intervention AUVs: the next challenge," *Annual Reviews in Control*, vol. 40, pp. 227–241, 2015.

- [2] L. Nurahmi, B. Pramujati, S. Caro et al., "Dimension synthesis of suspended eight cables-driven parallel robot for search-and-rescue operation," in *Proceedings of the 2017 International Conference on Advanced Mechatronics, Intelligent Manufacture, and Industrial Automation*, Surabaya, Indonesia, October 2017.
- [3] E. Cataldi and G. Antonelli, "Basic interaction operations for an underwater vehicle-manipulator system," in *Proceedings of the 2015. International Conference on Advanced Robotics*, Istanbul, Turkey, July 2015.
- [4] K. Li, Y. Zhang, and Q. Hu, "Dynamic modelling and control of a tendon-actuated lightweight space manipulator," *Aerospace Science and Technology*, vol. 84, pp. 1150–1163, 2019.
- [5] G. Loianno, Y. Mulgaonkar, C. Brunner et al., *Smartphones Power Flying robots*, Institute of Electrical and Electronics Engineers Inc., Hamburg, Germany, 2015.
- [6] K. Andersson, I. M. Militärvetenskapliga, A. M. Militärtekniska et al., "Modeling the impact of surface emissivity on the military utility of attack aircraft," *Aerospace Science and Technology*, vol. 65, pp. 133–140, 2017.
- [7] J. Thomas, G. Loianno, J. Polin, K. Sreenath, and V. Kumar, "Toward autonomous avian-inspired grasping for micro aerial vehicles," *Bioinspiration & Biomimetics*, vol. 9, no. 2, pp. 025010–025015, 2014.
- [8] M. Orsag, C. M. Korpela, S. Bogdan, and P. Y. Oh, "Hybrid adaptive control for aerial manipulation," *Journal of Intelligent & Robotic Systems*, vol. 73, no. 1–4, pp. 693–707, 2014.
- [9] R. Naldi, L. Gentili, and L. Marconi, "Modeling and control of the interaction between flying robots and the environment," *IFAC Proceedings Volumes*, vol. 43, no. 14, pp. 975–980, 2010.
- [10] A. E. Jimenez-Cano, J. Braga, G. Heredia et al., "Aerial manipulator for structure inspection by contact from the underside," in *Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, Hamburg, Germany, September 2015.
- [11] K. Alexis, G. Darivianakis, M. Burri, and R. Siegwart, "Aerial robotic contact-based inspection: planning and control," *Autonomous Robots*, vol. 40, no. 4, pp. 631–655, 2016.
- [12] P. E. I. Pounds, D. R. Bersak, and A. M. Dollar, "Grasping from the air: hovering capture and load stability," in *Proceedings of the IEEE International Conference on Robotics and Automation*, Shanghai, China, May 2011.
- [13] C. Korpela, P. Brahmabhatt, M. Orsag et al., "Towards the realization of mobile manipulating unmanned aerial vehicles (MM-UAV): peg-in-hole insertion tasks," in *Proceedings of the IEEE Conference on Technologies for Practical Robot Applications (TePRA)*, Woburn, MA, USA, April 2013.
- [14] C. Shiu, C. Hsu, and J. Chou, "Analysis of a quadrotor with a two-degree-of-freedom robotic arm," in *Proceedings of the 2016 International Symposium on Computer, Consumer and Control (IS3C)*, Xi'an, China, July 2016.
- [15] A. Suarez, G. Heredia, and A. Ollero, "Design of an anthropomorphic, compliant, and lightweight dual arm for aerial manipulation," *IEEE Access*, vol. 6, pp. 29173–29189, 2018.
- [16] G. Garimella and M. Kobilarov, "Towards model-predictive control for aerial pick-and-place," in *Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA)*, Seattle, WA, USA, May 2015.
- [17] E. N. Johnson and M. A. Turbe, "Modeling, control, and flight testing of a small-ducted fan aircraft," *Journal of Guidance, Control, and Dynamics*, vol. 29, no. 4, pp. 769–779, 2006.
- [18] R. Naldi, A. Torre, and L. Marconi, "Robust control of a miniature ducted-fan aerial robot for blind navigation in unknown populated environments," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 1, pp. 64–79, 2015.
- [19] G. Heredia, A. E. Jimenez-Cano, I. Sanchez et al., "Control of a multirotor outdoor aerial manipulator," in *Proceedings of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Chicago, IL, USA, September 2014.
- [20] F. Huber, K. Kondak, K. Krieger et al., *First Analysis and Experiments in Aerial Manipulation Using Fully Actuated Redundant Robot Arm*, IEEE, Piscataway, NJ, USA, 2013.
- [21] K. Kondak, F. Huber, M. Schwarzbach et al., "Aerial manipulation robot composed of an autonomous helicopter and a 7 degrees of freedom industrial manipulator," in *Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA)*, Hong Kong, China, May 2014.
- [22] M. Orsag, C. Korpela, and P. Oh, "Modeling and control of MM-UAV: mobile manipulating unmanned aerial vehicle," *Journal of Intelligent & Robotic Systems*, vol. 69, no. 1–4, pp. 227–240, 2013.
- [23] B. Yang, Y. He, J. Han, and G. Liu, "Rotor-Flying manipulator: modeling, analysis, and control," *Mathematical Problems in Engineering*, vol. 2014, Article ID 492965, 13 pages, 2014.
- [24] G. Chmaj, T. Buratowski, T. Uhl et al., "The dynamics influence of the attached manipulator on unmanned aerial vehicle," pp. 109–119, 2013.
- [25] D. Mellinger, Q. Lindsey, M. Shomin et al., "Design, modeling, estimation and control for aerial grasping and manipulation," in *Proceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Piscataway, NJ, USA, September 2011.
- [26] A. E. Jimenez-Cano, J. Martin, G. Heredia et al., "Control of an aerial robot with multi-link arm for assembly tasks," in *Proceedings of the 2013 IEEE International Conference on Robotics and Automation*, Karlsruhe, Germany, May 2013.
- [27] V. Lippiello and F. Ruggiero, "Cartesian impedance control of a UAV with a robotic arm," *IFAC Proceedings*, vol. 45, no. 22, pp. 704–709, 2012.
- [28] W. J. McCroskey and National Aeronautics And Space Administration Moffett Field Ca Ames Researchcenter, *A Critical Assessment of Wind Tunnel Results for the NACA 0012 Airfoil*, National Aeronautics And Space Administration Moffett Field Ca Ames Researchcenter, Moffett Field, CA, USA, 1987.
- [29] J. J. Craig, *Introduction to Robotics: Mechanics and control*, Pearson Prentice Hall, London, UK, 2005.
- [30] Y. Sun, J. Xu, H. Qiang, and G. Lin, "Adaptive neural-fuzzy robust position control scheme for maglev train systems with experimental verification," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 11, pp. 8589–8599, 2019.
- [31] Y. Sun, J. Xu, H. Qiang, C. Chen, and G. Lin, "Adaptive sliding mode control of maglev system based on RBF neural network minimum parameter learning method," *Measurement*, vol. 141, pp. 217–226, 2019.
- [32] J.-M. Pflimlin, P. Binetti, P. Souères, T. Hamel, and D. Trouchet, "Modeling and attitude control analysis of a ducted-fan micro aerial vehicle," *Control Engineering Practice*, vol. 18, no. 3, pp. 209–218, 2010.
- [33] P. V. Kimon and J. V. George, *Handbook of Unmanned Aerial Vehicles*, Springer, Cham, Switzerland, 2015.
- [34] I. B. Tijani, R. Akmeliawati, A. Legowo, and A. Budiyo, "Nonlinear identification of a small scale unmanned helicopter using optimized NARX network with multiobjective differential evolution," *Engineering Applications of Artificial Intelligence*, vol. 33, pp. 99–115, 2014.

- [35] J. F. Horn, *Aircraft and Rotorcraft System Identification: Engineering Methods with Flight Test Examples*, IEEE, New York, NY, USA, 2008.
- [36] P. Gahinet and P. Apkarian, "Frequency-domain tuning of fixed-structure control systems," in *Proceedings of 2012 UKACC International Conference on Control*, IEEE, Cardiff, UK, September 2012.
- [37] P. Apkarian, M. N. Dao, and D. Noll, "Parametric robust structured control design," *IEEE Transactions on Automatic Control*, vol. 60, no. 7, pp. 1857–1869, 2015.
- [38] N. Hovakimyan and C. Cao, *L1 Adaptive Control Theory: Guaranteed Robustness with Fast adaptation*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2010.
- [39] J.-B. Pomet and L. Praly, "Adaptive nonlinear regulation: estimation from the Lyapunov equation," *IEEE Transactions on Automatic Control*, vol. 37, no. 6, pp. 729–740, 1992.
- [40] S. Gage, "Creating a unified graphical wind turbulence model from multiple specifications," in *Proceedings of the AIAA Modeling and Simulation Technologies Conference and Exhibit*, Austin, TX, USA, 2003.

## Research Article

# Research on the Niche Evolution Game of Ecological Community Innovation of Corporate Venture Capital Based on Logistic Extended Complexity Model

Fanglin Meng <sup>1,2</sup>, Zengrui Tian <sup>2</sup>, Beiquan Chang <sup>3</sup>, Hongxin Yu <sup>4</sup>, and Shuai Zhang<sup>5</sup>

<sup>1</sup>Business School, Sanda University, Shanghai 201209, China

<sup>2</sup>Glorious Sun School of Business and Management, Donghua University, Shanghai 200051, China

<sup>3</sup>School of Public Policy & Management, Tsinghua University, Beijing 100084, China

<sup>4</sup>Business Economics College, Shanghai Business School, Shanghai 200235, China

<sup>5</sup>University of Greenwich, London SE10 9LS, UK

Correspondence should be addressed to Zengrui Tian; [flmeng@sandau.edu.cn](mailto:flmeng@sandau.edu.cn)

Received 17 April 2020; Revised 26 May 2020; Accepted 3 June 2020; Published 28 August 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Fanglin Meng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the gestation and development of new technologies, new products, new formats, and new models, venture capital investment, as one of the most important forms of open innovation in large companies, plays an increasingly important role in the innovation of mature large companies and entrepreneurial enterprises. To deal with the complex and dynamic environment, the niche of Corporate Venture Capital (CVC) ecological community is investigated from the perspective of the innovation ecosystem. By analyzing the innovation of CVC ecological community with the use of the logistic expansion model, this paper analyzes the stability of evolution game through the replicator dynamic equation and discusses ten parameters of niche state. In the end, we conclude that there are four optimization strategies in the coevolution of major corporations and entrepreneurial firms, namely, niche separation, niche expansion, niche K-R, and niche alliance.

## 1. Introduction

Innovation has become the endogenous driving force and core competitive advantage of a country's economic growth. In the complex and dynamic digital economy era, it is difficult for enterprises to maintain traditional competition barriers. To keep growing continuously, enterprises must deal with issues such as the increasingly uncertain environment, challenging value creation, and oversupply choices for customers. The traditional value chain creation model is changing to the shared destiny community with a symbiosis creation model. Therefore, large companies are constantly seeking the path of open empowerment. Corporate Venture Capital (CVC) has become an important model and tool for open innovation, access to external technology sources, and value creation. The global venture capital funding in 2018 reached 254 billion US dollars, and the financing amount in

2017 was 174 billion RMB, of which the global company's total venture capital investment was 312 billion US dollars. Investment activities increased by 19% over 2016, and total investment increased by 18% [1]. Companies in countries like Britain, China, and India have hit record highs in their venture capital. The global CVC investment industry has strong interests in areas with high technological innovation and fierce competitions such as Internet, healthcare, and mobile technology.

The company's venture capital originated in the United States in the 1960s. It is now one of the main ways for mature large companies to develop outwards. It has a strategic appeal because it provides multiple choice platforms for large companies to innovate. Corporate innovation requires venture capital as a mechanism and driving force. Corporate Venture Capital aims to achieve open innovation, obtain the latest technology in related fields, and overcome the soft



constraints of the company's internal R&D budget so that they can improve technology innovation efficiency. It could also reduce the risk of investment in independent R&D and improve organizational capabilities to foster an innovation culture. It benefits the corporate by realizing the value of strategic mergers and acquisitions and initiating entrepreneurial investment in strategic innovation projects. As company's venture capital has a high degree of risk tolerance and a long investment cycle, it can invest in the startups that are struggling to obtain traditional independent venture capital (IVC). It promotes and optimizes mutual regeneration and achieves a better evolutionary cycle for both parties. While mature companies deeply cultivate the Red Sea, they must also actively explore in the Blue Ocean. CVC is a hybrid model that stimulates innovation. It expands industrial boundaries, rebuilds market boundaries, explores cutting-edge technologies, encourages open innovation, and expands organizational structures for large companies. It is regarded as an effective paradigm to promote the evolution of the industry ecosystem. It is also an important way for large companies to acquire innovative technologies and industrial integration [2]. CVC activities are a "dual and two-way" value creation process that includes both strategic and financial values, while creating value for mature large companies and startups. With the advent of the digital economy 2.0 era, ecological collaborative innovation models have gradually taken shape. Typical digital economy companies in China, such as Alibaba, Tencent, Xiaomi, and Baidu, as well as traditional enterprises such as Haier and Fosun, are actively building open innovation enterprise ecosystems. They integrate the company's own technology, capital, and market advantages with external resources to develop new technologies, new products, and new models for new markets.

## 2. Related Work

Hannah and Eisenhardt proposed that, in addition to the relationship between competition and cooperation, enterprises also need to continuously meet customer needs through innovation [3]. Enterprises have evolved from the role of an individual player in the industry into part of the industrial ecosystem. Luo and Ratchford took Apple, IBM, Ford, and Wal-Mart as examples to study focal companies which develop service, technology, and value network platforms to build their own unique business ecosystems and obtain value returns [4]. Yao and Zhou used the theory of natural ecosystem evolution to study the innovation path of high-tech enterprises and found that the dependence of enterprise innovation paths has ecological genetic and variability characteristics [5]. Daniela et al. in "Evolutionary Theory of Economic Change" explain economic changes from a perspective of dynamic evolution. Evolutionary game theory pays attention to the change of population structure and uses evolutionary stable strategy (ESS) to represent a stable state that can resist the invasion of mutation strategy [6]. Levinthal proposed that corporate adaptation and environmental selection are the main paths of population evolution [7]. Zaman et al. believe that internal and external

resources of an enterprise are equally important. Enterprises use external resources and external channels to help commercialize their new technological achievements. They also emphasize that enterprises should quickly implement innovation, reduce innovation costs, and work with professional venture capital institutions to jointly improve innovation performance [8].

One of the most challenging 125 scientific issues in the 21st century in science is, "how does cooperative behavior evolve?" The innovation ecosystem has the same characteristics of natural evolution, integration, self-organization, periodicity, and openness as natural ecosystems and shares the same characteristics in power, genetics, evolution, and feedback. Important scholars in the CVC field, Jog and Mcconomy, creatively proposed that mature large companies have regarded CVC not only as the window for technology discovery but also as an ecosystem around large companies [9]. This conclusion is based on previous research on CVC enterprise value creation, CVC investment enterprise entrepreneurial performance, investment conditions, and cross-organizational knowledge acquisition. Based on the analysis of nearly 300 CVCs, Jog and Mcconomy proposed that the increasing number of CVCs would contribute in better short-term and long-term performance of the ecosystem, which extends a previous view shared by scholars that the CVC is only an incubator-level participant. Instead, as an ecosystem strategy, CVC intends to "build a constellation" [10]. Therefore, from the perspective of the ecosystem and the niche theory, it is proposed that the CVC ecological community, formed by mature large companies and startups in the form of CVC projects, achieves synchronization through interactions between CVC functions, behavioral processes, and the external environment. In the symbiotic mode, a symbiotic evolution can be developed with the growth of large companies and entrepreneurial enterprises.

## 3. CVC Niche for Major Corporations and Entrepreneurial Firms

In 1838, the Dutch mathematical biologist Verhulst proposed a logistic equation to study the growth of biological populations. It was found that the population grew fastest at the beginning. When it grew to a certain value, the speed began to slow down until it finally decreased to zero (i.e., stop growing). In 1900, Italian mathematician Volterra proposed the predator and prey populations model. In 1925, Lotka proposed mathematical ecology in chemical reactions. Then in 1926, Volterra used the Lotka-Volterra model to demonstrate the rule of fish population in the port of Fiume. Lotka-Volterra model, which evolved from the logistic extended model, can better explain the relationship between large companies and startups in the CVC ecological community. This model, as an evolutionary game for species or population formation, provides a new perspective to understand community structure optimization and niche evolution [11].

The growth of the corporate ecological community involved in CVC also generally conforms to the mechanism of

the logistic development; that is, the development of the CVC ecological community is relatively immature: the scale is not large; the growth space is large; and the development speed is fast at S1. However, at this stage, for large enterprises and startups, the initial capital investment, human investment, and technical investment are relatively large. The symbiotic subjects of the CVC ecological community are still in the running-in period with high risks. At node P1, either CVC is offside to enhance the niche, occupying a favorable position in the innovation and entrepreneurship ecosystem, or it may gradually decline. CVCs that operate smoothly, realize technological innovation, and have shared and coordinated development can successfully enter the second stage of the ecological niche beyond P1. In the second stage of S2, the ecological niche of CVC continued to expand; the benefits continued to increase; innovation results and innovative technologies were rapidly transformed; large companies achieved their strategic goals and financial goals; and startup companies also achieved growth. At the second branch point P2, with great innovation potential, the CVC ecological community, formed by large companies with startups, integrates more resources and continues to develop. The symbiotic system will rise to a higher stage, while the ecological shrinkage may also occur. Figure 1 is the trend of CVC ecological community evolution.

Therefore, the evolution of the niche also reflects the characteristics of repeated games between large companies and startups in the CVC ecological community. CVC activities are also innovative activities. Large companies form CVC ecological communities with startups through venture capital for technological innovation. They aim to realize value cultivation and value innovation and continuously develop new competitive advantages in the industry competition. By improving their own niche, they can achieve their strategic goals and financial goals.

From the perspective of biology and evolutionary games, the development of the CVC ecological community is constrained by various factors such as talent, technology, capital, and services, in specific time and space dimensions. The ecological factor domain  $C_i = f(z_1, z_2, z_3, \dots, z_n)$  is used to represent the various factors  $z_1, z_2, z_3, \dots, z_n$  that indicate the constraints for the development of the CVC ecological community. There is a range value  $N_i$  of the niche width  $W_i$  of ecological community  $i$ , which represents the competition coefficient  $\alpha_i$  of the CVC ecological community.  $\beta$  represents the niche overlap value of the CVC ecological community, and  $\chi_i$  indicates the impact of CVC investment strategies such as combination strategies, capital injection strategies, exit methods, and space preferences on the innovation of the CVC ecological community.  $S_i$  represents the niche status (i) of the CVC ecological community at time of  $t$ , and  $dS(t)/dt$  represents the rate of change in the niche width for large enterprises and entrepreneurial enterprises in the CVC ecology community. Adjusted from the Lotka-Volterra model, the niche evolution equations of the major corporations (MC) and entrepreneurial firms (EF) in the CVC ecosystem are set as follows:

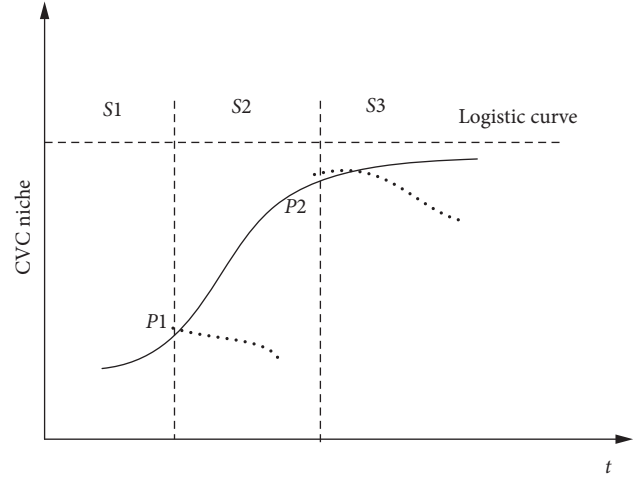


FIGURE 1: The trend of CVC ecological community evolution.

$$\begin{cases} \frac{dS_1(t)}{dt} = r_1 S_1 (\rho_1 + \chi_1 - \alpha_1 - S_1 - \beta S_2), \\ \frac{dS_2(t)}{dt} = r_2 S_2 (\rho_2 + \chi_2 - \alpha_2 - S_2 - \beta S_1). \end{cases} \quad (1)$$

Here,  $S_1$  and  $S_2$  represent the niche width limit of MC and EF;  $\rho_1$  and  $\rho_2$  represent niche saturation coefficient of MC and EF in the CVC ecological community;  $\alpha_1$  and  $\alpha_2$  represent competition coefficients:  $\alpha_1$  is the competitive effect of EF on MC, while  $\alpha_2$  is the competitive effect of MC on EF; and  $\beta$  represents niche overlap between MC and EF in the CVC ecological community.

When  $\beta = 1$ , there is a complete niche overlap between MC and EF. When  $\beta = 0$ , there is a complete niche separation for MC and EF. If the value of  $\beta$  is between 0 and 1, there is partial overlap between MC and EF. The  $\beta$  is proportional to  $\alpha_1$  and  $\alpha_2$ , which means the bigger the overlap between MC and EF, the more the fierce competition.

According to the above differences in  $\beta$ , the evolution of the niche of large companies and startups in the CVC ecological community is discussed as follows [12]:

- (1) When  $\beta = 1$ , there is a complete niche overlap between MC and EF.

If  $\begin{cases} dS_1(t)/dt = 0 \\ dS_2(t)/dt = 0 \end{cases}$ , the balanced position includes

$$H_1(\rho_1 + \chi_1 - \alpha_1, 0), \quad H_2(\rho_2 + \chi_2 - \alpha_2, 0), \quad H_3(0, \rho_1 + \chi_1 - \alpha_1), \quad H_4(0, \rho_2 + \chi_2 - \alpha_2), \quad O(0, 0).$$

The equation for MC in the CVC ecological community is  $\rho_1 + \chi_1 - \alpha_1 - S_1 - S_2 = 0$ .

The equation for EF in the CVC ecological community is  $\rho_2 + \chi_2 - \alpha_2 - S_2 - S_1 = 0$ .

As there is an equal slope, the position of the two parallel straight lines  $L_1, L_2$  depends on the values of parameters  $\rho_1 + \chi_1 - \alpha_1$  and  $\rho_2 + \chi_2 - \alpha_2$ .

When  $\rho_1 + \chi_1 - \alpha_1 > \rho_2 + \chi_2 - \alpha_2$ , we can obtain Figure 2 in the following.

The niche width of the MC in the CVC ecological community is larger than the niche width of the EF.

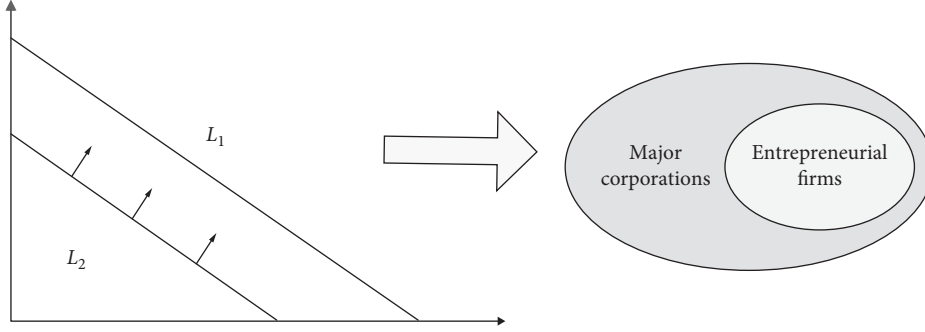


FIGURE 2: Evolution of niche overlap, when MC contains EF.

MC completely includes the EF, indicating that the MC occupies the entire niche space.

When  $\rho_2 + \chi_2 - \alpha_2 > \rho_1 + \chi_1 - \alpha_1$ , we can get Figure 3 in the following.

Entrepreneurial firm's niche width ( $L_2$ ) in the CVC ecological community is larger than that of major corporation ( $L_1$ ). EF completely includes MC. This indicates that EF occupies the entire niche space. The result of evolution will also approach EF. Therefore, when the MC and EF in the CVC ecological community completely overlap, the direction of the niche evolution depends on the technologies, capital, human resources, and innovation resources used by the MC and EF. The one occupying the resource advantage will eventually control the evolution.

- (2) When  $\beta = 0$ , there is a complete niche separation for MC and EF.

If  $\begin{cases} dS_1(t)/dt = 0 \\ dS_2(t)/dt = 0 \end{cases}$ , the balanced position includes

$$H_1(\rho_1 + \chi_1 - \alpha_1, \rho_2 + \chi_2 - \alpha_2),$$

$$H_2(\rho_1 + \chi_1 - \alpha_1, 0), H_3(0, \rho_2 + \chi_2 - \alpha_2), O(0, 0).$$

The equation for MC in the CVC ecological community is  $L_1: S_1 = \rho_1 + \chi_1 - \alpha_1$ .

The equation for EF in the CVC ecological community is  $L_2: S_2 = \rho_2 + \chi_2 - \alpha_2$ .

We can obtain a separation as shown in Figure 4.

$L_1$  and  $L_2$  divide four areas in the quadrant. In the areas below  $L_1$  ( $dS_1(t)/dt < 0$ ), the evolution direction rises, while in the areas above  $L_1$  ( $dS_1(t)/dt > 0$ ), the evolution direction goes downward. In the areas on the left of  $L_2$  ( $dS_2(t)/dt < 0$ ), the evolution direction goes right, while in the areas on the right of  $L_2$  ( $dS_2(t)/dt > 0$ ), the evolution direction goes left.

In quadrant 1,  $dS_1(t)/dt > 0$  and  $dS_2(t)/dt < 0$ .

In quadrant 2,  $dS_1(t)/dt > 0$  and  $dS_2(t)/dt > 0$ .

In quadrant 3,  $dS_1(t)/dt < 0$  and  $dS_2(t)/dt < 0$ .

In quadrant 4,  $dS_1(t)/dt < 0$  and  $dS_2(t)/dt > 0$ .

Therefore, we can conclude that the result of evolution tends to equilibrium  $H_1(\rho_1 + \chi_1 - \alpha_1, \rho_2 +$

$\chi_2 - \alpha_2)$ , where the large companies and startups in the CVC ecological community occupy their respective niche. Its niche size is related to the niche saturation coefficients  $\rho_i$  of both parties, the competition coefficient  $\alpha_i$  of large enterprises and startups in the CVC ecosystem, and the impact coefficient  $\chi_i$  of CVC ecological community innovation CVC investment strategy (CVC portfolio strategy, CVC investment stage strategy, CVC exit method, CVC investment space preference).

- (3) When  $0 < \beta < 1$ , MC and EF niche partially overlap. The evolution path depends on the ecological domain space of their respective resources.

If  $\begin{cases} dS_1(t)/dt = 0 \\ dS_2(t)/dt = 0 \end{cases}$ , the balanced position includes

$$O(0, 0), H_1(\rho_1 + \chi_1 - \alpha_1, 0),$$

$$H_2(0, \rho_1 + \chi_1 - \alpha_1/\beta), H_3(\rho_2 + \chi_2 - \alpha_2/\beta, 0), \text{ and } H_4(0, \rho_2 + \chi_2 - \alpha_2):$$

$$G\left(\frac{\rho_1 + \chi_1 - \alpha_1 - \beta(\rho_2 + \chi_2 - \alpha_2)}{1 - \beta^2}, \frac{\rho_2 + \chi_2 - \alpha_2 - \beta(\rho_1 + \chi_1 - \alpha_1)}{1 - \beta^2}\right). \quad (2)$$

As the values of the parameters are different, the analytical expressions have different evolution trends. There are four competition situations between the two populations:

*Situation 1.* When  $(\rho_1 + \chi_1 - \alpha_1/\beta) > \rho_2 + \chi_2 - \alpha_2$  and  $(\rho_2 + \chi_2 - \alpha_2/\beta) < \rho_1 + \chi_1 - \alpha_1$ .

In the area surrounded by the quads  $H_1, H_2, H_3, H_4$  (Figure 5), where  $dS_1(t)/dt < 0, dS_2(t)/dt > 0$ , large companies have not reached the maximum capacity for growth; i.e., there is still room for growth and development. While, startups have reached the maximum capacity; i.e., there is little room for continued growth. Large companies have mastered related technological innovations. The corresponding evolution is shown in Figure 5.

*Situation 2.* When  $(\rho_1 + \chi_1 - \alpha_1/\beta) < \rho_2 + \chi_2 - \alpha_2$  and  $(\rho_2 + \chi_2 - \alpha_2/\beta) > \rho_1 + \chi_1 - \alpha_1$ .

In the area surrounded by the quads  $H_1, H_2, H_3, H_4$  (Figure 6), where  $dS_1(t)/dt > 0, dS_2(t)/dt > 0$ ,

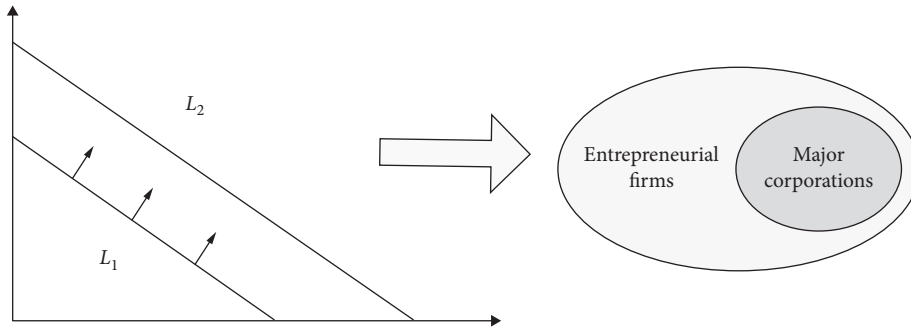


FIGURE 3: Evolution of niche overlap, when EF contains MC.

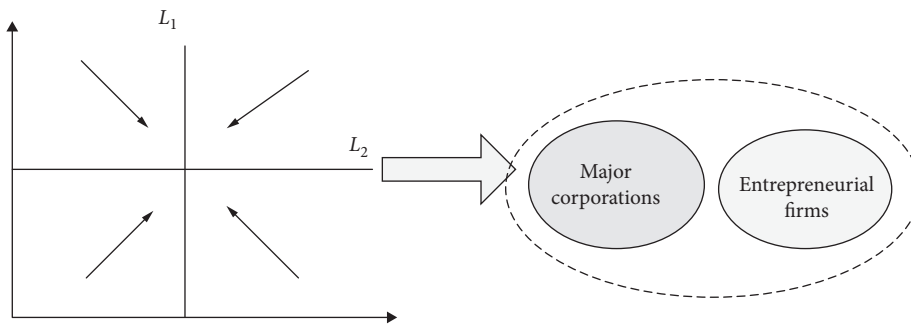


FIGURE 4: Niche separation of CVC.

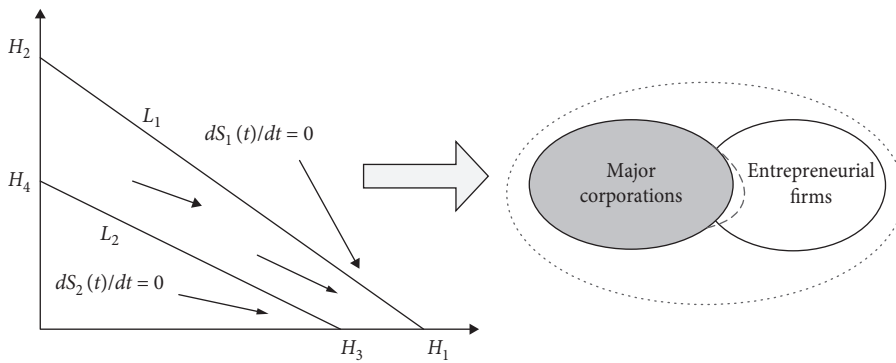


FIGURE 5: Large companies occupy part of the startup's niche.

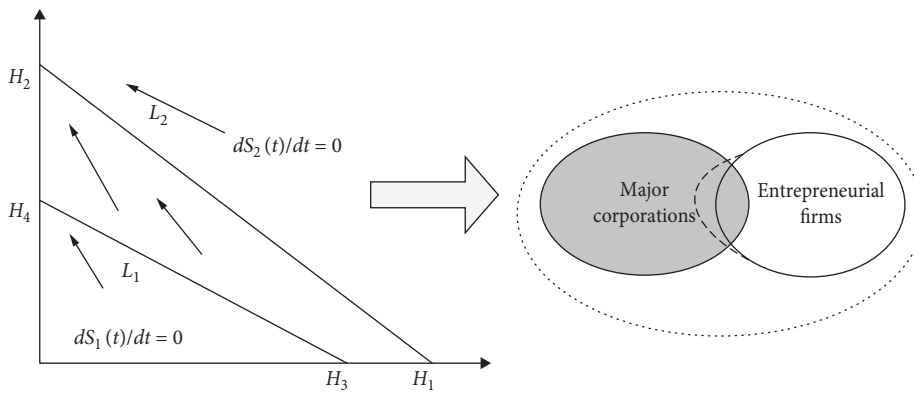


FIGURE 6: Startups occupy the ecological domain of large companies.

entrepreneurial enterprises in the CVC ecosystem can continue to develop, but large companies have reached their limits. Therefore, entrepreneurial enterprises can invade some ecological domains for large companies, as shown in Figure 6.

*Situation 3.* When  $(\rho_2 + \chi_2 - \alpha_2/\beta) > \rho_1 + \chi_1 - \alpha_1$  and  $(\rho_1 + \chi_1 - \alpha_1/\beta) > \rho_2 + \chi_2 - \alpha_2$ ,  $L_1$  and  $L_2$  come across at point G, as shown in Figure 7.

$$G\left(\frac{\rho_1 + \chi_1 - \alpha_1 - \beta(\rho_2 + \chi_2 - \alpha_2)}{1 - \beta^2}, \frac{\rho_2 + \chi_2 - \alpha_2 - \beta(\rho_1 + \chi_1 - \alpha_1)}{1 - \beta^2}\right). \quad (3)$$

In the area of H2GH4, there is still room for growth of startups, while large companies have developed to the maximum capacity, approaching G. In the GH1H3 region, large companies continue to grow, while the growth of startups has reached the upper limit. The movement trends of both parties will gradually approach point G, where G is the equilibrium point for the coexistence of large enterprises and entrepreneurial enterprises in the CVC ecological community.

*Situation 4.* When  $(\rho_2 + \chi_2 - \alpha_2/\beta) < \rho_1 + \chi_1 - \alpha_1$  and  $(\rho_1 + \chi_1 - \alpha_1/\beta) < \rho_2 + \chi_2 - \alpha_2$ .

$L_1$  and  $L_2$  come across at point P, as shown in Figure 8.

In the H2PH4 and H1PH3 areas, both large companies and startups have room for continued innovation and growth. Both sides may have the opportunity to win in the competition process, so they have not reached an equilibrium state. The niche occupation depends on what the two parties have in the initial ecological domain. The larger the initial ecological domain is, the more resources it possesses, and the easier it is to win in the CVC ecological community. This also makes CVC ultimately achieve the strategic goals and financial goals of large companies by IPO, mergers, and acquisitions. While startups get mature, big companies can quit.

In the above analysis, situations (1), (2), and (4) are all unbalanced and stable. In the balanced state (3), large enterprises and startups in the CVC ecosystem can coexist and maximize their own added value.

#### 4. Construction of the Evolutionary Game Model for the CVC Ecological Community Innovation

Evolutionary game theory exists in Darwin's idea of natural selection. Based on the dual theory of game theory and biological evolution, this section analyzes the formation mechanism of interaction between populations in ecological communities and studies population growth, evolution, and stability [13].

*4.1. Evolutionary Game Analysis.* The symbiotic evolution between large companies and startups in the CVC ecological community will increase the adaptability of both parties and thus their respective niche. The evolutionary game theory has been used to analyze the operating mechanism of large

companies and startups. Here we use evolution stability strategy to analyze the evolution path of CVC ecological community. The strategy adjustment process, trends, and stability of large players and startups in priority game players are considered. ESS evolutionary stability strategy and Taylor and Jonker's replicator dynamic equations are used as well [14]. The niche evolution process of CVC ecological community innovation also requires mutation and selection mechanisms [15]. The formation and development of the CVC eco-community is the game behavior between large companies, startups, and other related entities. Its formation and development are also the result of the dynamic evolution of multiparty games. When there is a two-way causal relationship between a large company and a startup company under a certain feedback mechanism, the adaptive change of the large company will change the adaptability of the startup company, while the change of the startup company will further affect the change of the large company. By assessing their participation in CVC investment activities, the two parties will promote the continuous evolution of the CVC ecological community when the results of multiple rounds of game play with other entities are greater than their nonparticipation.

#### 4.2. Research Hypothesis and Model Building

*4.2.1. Research Hypothesis.* The changes in the niche of large companies and startups in the CVC ecosystem are affected by a variety of ecological factors in the symbiotic environment of venture capital. At the same time, the strategy obtained by the opponent based on the irrational selection is an important factor. Entrepreneurial enterprises will play dynamic games. First, assuming that the strategy set of MC is niche maintenance and niche expansion, the strategy set of an entrepreneurial company can be participation in competition and acceptance. In this case, there is a two-stage game in the CVC investment activity process. This can be regarded as an evolutionary game of dual population and dual strategy [16]. Second, assuming that both large companies and startups are bounded in rationality, they choose the best strategy based on their own situation and environment. Third, in the CVC ecological community, large companies, and startups make strategic adjustments based on their strategy and the external environment. A repeated game would happen as the adjustment also depends on the choices and the performance of these companies.

*4.2.2. Construction of Fitness Function Matrix.* Assuming that the niche separation of large companies and startups does not affect them, the two types of corporate populations in their own ecological space and time have the benefits of independent innovation  $R_1$  and  $R_2$ . When the MC population chooses to expand its niche through CVC activities, it can realize innovation benefits, including the benefits created by knowledge sharing as a result of sharing various resources and technology spillovers. It can be represented by  $\lambda(a + b)\Delta r$ , where  $\lambda$  is the absorptive capacity coefficient of the large company population;  $a$  and  $b$  are the knowledge

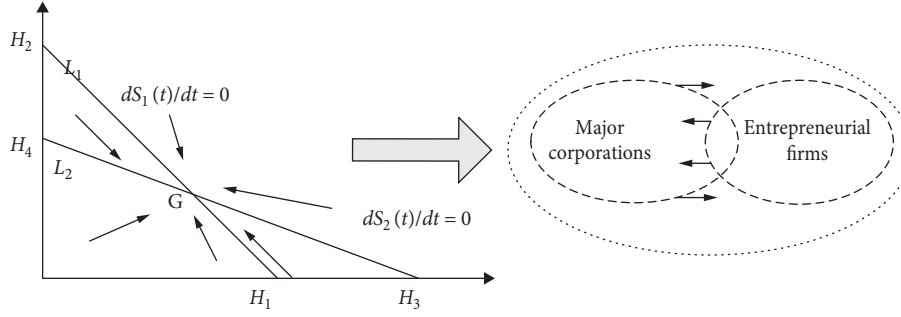


FIGURE 7: Ecological stability of large companies and startups.

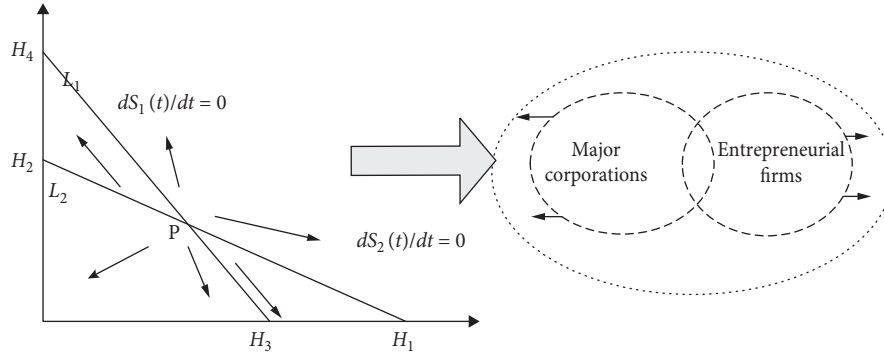


FIGURE 8: Large companies and startups continue to grow and cannot reach balance.

sharing coefficient and the technology spillover coefficient, respectively; and  $\Delta r$  is unit excesses benefit. If the MC expands, the EF also carries out the niche expansion, which can achieve excess returns. Both parties need to pay a certain cost during the niche expansion process. The cost paid by the large company can be expressed as  $C_1$ , and the cost paid by startup company is represented by  $C_2$ . There are also large companies that expand their niche, and startups choose to keep the same niche. In this case, they can get a profit  $e$ . The probability that a large company will expand its niche is  $p$ , and the probability of maintaining it is  $1 - p$ . The probability of participating in competition based on feedback from startups is  $q$ , and the probability of maintaining the original ecology is  $1 - q$ . The game fitness payment matrix is established in Table 1.

#### 4.3. Evolution Path Analysis

**4.3.1. Replicator Dynamic Equation Solving.** According to the game fitness function matrix above, it can be concluded that, for the MC, which chooses to expand the niche, that is, to carry out CVC investment activities, the expected return is

$$\begin{aligned} U_{11} &= q(R_1 + \lambda(a+b)\Delta r - C_1) + (1-q)(R_1 - C_1) \\ &= R_1 + q\lambda(a+b)\Delta r - C_1, \\ U_{12} &= q(R_1 + e) + (1-q)R_1 = qe + R_1, \end{aligned} \quad (4)$$

$$\begin{aligned} \bar{U}_1 &= pU_{11} + (1-p)U_{12} \\ &= p(q\lambda(a+b)\Delta r - qe - C_1) + qe + R_1. \end{aligned}$$

TABLE 1: Niche game fitness function matrix.

| Subject niche in CVC |             | EF   |                          |
|----------------------|-------------|--|--------------------------|
|                      |             | Expanding niche  | Maintaining niche        |
| MC                   | Expanding   | $R_1 + \lambda(a+b)\Delta r - C_1$<br>$R_2 + [1 - \lambda(a+b)]\Delta r - C_2$ | $R_1 - C_1$<br>$R_2 + e$ |
|                      | Maintaining | $R_1 + e$<br>$R_2 - C_2$   | $R_1$<br>$R_2$           |

In the same way, for the EF, the expected income and expansion of the niche expansion are  $U_{21}$  and  $U_{22}$ , while the average expectation of the startup enterprise is  $\bar{U}_2$ :

$$\begin{aligned} U_{21} &= p[R_2 + [1 - \lambda(a+b)]\Delta r - C_2] + (1-p)(R_2 - C_2) \\ &= R_2 + p[1 - \lambda(a+b)]\Delta r - C_2, \\ U_{22} &= p(R_2 + e) + (1-p)R_2 = e + R_2, \\ \bar{U}_2 &= qU_{21} + (1-q)U_{22}, \\ &= q[R_2 + p[1 - \lambda(a+b)]\Delta r - C_2] + (1-q)[pe + R_2] \\ &= q[p[1 - \lambda(a+b)]\Delta r - pe - C_2] + pe + R_2. \end{aligned} \quad (5)$$

According to the replicator dynamic formula of the evolutionary game, the replicator dynamic equations of large companies and startups in the CVC ecosystem can be obtained as follows:

$$\begin{aligned}
F(p) &= \frac{dp}{dt} = p(U_{11} - \overline{U}_1) \\
&= p(1-p)[q(\lambda(a+b)\Delta r - e) - C_1], \\
F(q) &= \frac{dq}{dt} = q(U_{22} - \overline{U}_2) \\
&= q(1-q)\{p[[1 - \lambda(a+b)]\Delta r - e] - C_2\}.
\end{aligned} \tag{6}$$

4.3.2. *Evolutionary Stability Strategies for MC.* According to the dynamic equations for MC, we have

$$\begin{aligned}
F(p) &= \frac{dp}{dt} = p(U_{11} - \overline{U}_1) \\
&= p(1-p)[q(\lambda(a+b)\Delta r - e) - C_1].
\end{aligned} \tag{7}$$

Three stable points of the equation can be obtained, namely,  $p_1^* = 0$ ,  $p_2^* = 1$ , and  $p_3^* = C_1/\lambda(a+b)\Delta r - e$ . If an evolutionary stable strategy is required, the derivative dynamic equation needs to be derived and satisfy  $d^2p/d^2t < 0$ . We will obtain

$$\frac{d^2p}{d^2t} = (1-2p)[q(\lambda(a+b)\Delta r - e) - C_1]. \tag{8}$$

When  $q > C_1/\lambda(a+b)\Delta r - e$ , we get  $p_2^* = 1$ , an evolutionary expansion strategy, which means MC will choose niche expansion strategy. When  $q < C_1/\lambda(a+b)\Delta r - e$ , we get  $p_1^* = 0$ , an evolutionary stability strategy, which means MC will choose strategies to maintain their original niche. When  $q = C_1/\lambda(a+b)\Delta r - e$ ,  $dp/dt$  equals zero, which is always a stable status.

4.3.3. *Evolutionary Stability Strategies for EF.* According to dynamic equations for EF, we have

$$\begin{aligned}
F(q) &= \frac{dq}{dt} = q(U_{22} - \overline{U}_2) \\
&= q(1-q)\{p[[1 - \lambda(a+b)]\Delta r - e] - C_2\}.
\end{aligned} \tag{9}$$

Three stable points of the equation can be obtained, namely,  $q_1^* = 0$ ,  $q_2^* = 1$ , and  $q_3^* = C_2/[1 - \lambda(a+b)]\Delta r - e$ .

When  $p = C_2/[1 - \lambda(a+b)]\Delta r - e$ , we get  $dq/dt = 0$ , with  $q$  being in stable status. When  $p > C_2/[1 - \lambda(a+b)]\Delta r - e$ , we get  $dq/dt < 0$ , which is an evolutionary expansion strategy. EF also chooses to expand its niche to enter MC's related industries. When  $p < C_2/[1 - \lambda(a+b)]\Delta r - e$ , we get  $q_1^* = 0$ , which is an evolutionary stable strategy. EF maintains its original niche.

The above discussion does not consider the relationship and size of the parameters. When the combination of parameters changes, the game strategy of large enterprises and entrepreneurial enterprises in the CVC ecological community changes.

4.3.4. *Impact of Parameter Changes on Evolutionary Game Strategies.* The first topic is about the effect of parameter changes on the evolutionary stability strategy of MC:

- (1) For  $0 < C_1/\lambda(a+b)\Delta r - e < 1$ , evolutionary stable equilibrium is determined by two situations of  $q$ . When  $q > C_1/\lambda(a+b)\Delta r - e$ , we get  $p_2^* = 1$ , an evolutionary expansion strategy (EES), while when  $q < C_1/\lambda(a+b)\Delta r - e$ , we get  $p_1^* = 0$ , an evolutionary stable strategy (ESS).
- (2) For  $C_1/\lambda(a+b)\Delta r - e \geq 1$ , we always get  $q < C_1/\lambda(a+b)\Delta r - e$ , and  $p_1^* = 0$ , which is ESS. That is, the cost of niche expansion is high, and large companies maintain their original niche.

The second topic is about the impact of changes in parameters on the evolutionary stability strategy of EF:

- (1) For  $0 < C_2/[1 - \lambda(a+b)\Delta r - e] < 1$ , there are two possibilities. When  $p > C_2/[1 - \lambda(a+b)\Delta r - e]$ , we get  $q_2^* = 1$ , an EES. When  $p < C_2/[1 - \lambda(a+b)\Delta r - e]$ , we get  $q_1^* = 0$ , an ESS.
- (2) For  $C_2/[1 - \lambda(a+b)\Delta r - e] \geq 1$ , we always get  $p < C_2/[1 - \lambda(a+b)\Delta r - e]$ ,  $q_1^* = 0$ , which is an ESS.

By a comprehensive analysis of the stability strategies of the two sides of the game analyzed above, we get five balance points, namely,  $O(0, 0)$ ,  $P(1, 0)$ ,  $Q(0, 1)$ ,  $R(1, 1) = E(C_1/\lambda(a+b)\Delta r - e, C_2/[1 - \lambda(a+b)]\Delta r - e)$ .

#### 4.4. Game Process with Niche Parameters

4.4.1. *Evolution Strategy of MC in the CVC Ecological Community.* When considering the status and momentum of large companies and startups in the CVC ecosystem, that is, the growth of large companies' innovation efforts and potential innovation capabilities, the fitness function needs to consider the status value of the large company population (T1) and potential value (S1). The game player's entrepreneurial state value is T2, and the potential value is S2. The higher the values of T1, S1, T2, and S2, the stronger the ability to obtain innovative resources and the stronger the ability to expand the niche. The matrix above can be adjusted to Table 2.

The corresponding income formula is as follows:

$$\begin{aligned}
U_{11} &= q(R_1 + \lambda(a+b)\Delta r - C_1) + (1-q)(R_1 - C_1), \\
U_{12} &= q(R_1 + e + T_1S_1) + (1-q)R_1 = q(e + T_1S_1) + R_1, \\
\overline{U}_1 &= pU_{11} + (1-p)U_{12} \\
&= p(q\lambda(a+b)\Delta r - q(e + T_1S_1) - C_1) + q(e + T_1S_1) + R_1.
\end{aligned} \tag{10}$$

The dynamic equation is as follows:

$$\begin{aligned}
F(p) &= \frac{dp}{dt} = p(U_{11} - \overline{U}_1) \\
&= p(1-p)[q(\lambda(a+b)\Delta r - e + T_1S_1) - C_1].
\end{aligned} \tag{11}$$

TABLE 2: Population fitness function matrix in CVC ecological community.

| Subject niche<br>in CVC |             | EF   |                                   |
|-------------------------|-------------|--|-----------------------------------|
|                         |             | Expanding niche  | Maintaining<br>niche              |
| MC                      | Expanding   | $R_1 + \lambda(a+b)\Delta r - C_1$<br>$R_2 + [1 - \lambda(a+b)]\Delta r - C_2$ | $R_1 - C_1$<br>$R_2 + e + T_2S_2$ |
|                         | Maintaining | $R_1 + e + T_1S_1$<br>$R_2 - C_2$  | $R_1$<br>$R_2$                    |

For  $F(p) = 0$ , the stable points are  $q_1^* = 0$ ,  $q_2^* = 1$ , and  $q_3^* = C_2/[1 - \lambda(a+b)]\Delta r - (e + T_1S_1)$ :

- (1) When  $q = C_1/\lambda(a+b)\Delta r - (e + T_1S_1)$ ,  $dp/dt$  equals zero, and the value of  $p$  is stable
- (2) When  $q > C_1/\lambda(a+b)\Delta r - (e + T_1S_1)$ ,  $d^2p/d^2t < 0$ , we obtain  $p_2^* = 1$ , an EES
- (3) When  $q < C_1/\lambda(a+b)\Delta r - (e + T_1S_1)$ ,  $d^2p/d^2t > 0$ , we obtain  $p_1^* = 0$ , an ESS

4.4.2. *Evolutionary Strategies of EF.* The corresponding income formula is as follows:

$$\begin{aligned}
U_{21} &= p[R_2 + [1 - \lambda(a+b)]\Delta r - C_2] + (1-p)(R_2 - C_2) \\
&= R_2 + p[1 - \lambda(a+b)]\Delta r - C_2, \\
U_{22} &= p(R_2 + e + T_2S_2) + (1-p)R_2 = p(e + T_2S_2) + R_2, \\
\overline{U}_2 &= qU_{21} + (1-q)U_{22} \\
&= q[R_2 + p[1 - \lambda(a+b)]\Delta r - C_2] \\
&\quad + (1-q)[p(e + T_2S_2) + R_2] \\
&= q[p[1 - \lambda(a+b)]\Delta r - p(e + T_2S_2) - C_2] \\
&\quad + p(e + T_2S_2) + R_2.
\end{aligned} \tag{12}$$

The dynamic equation is as follows:

$$\begin{aligned}
F(q) &= \frac{dq}{dt} = q(U_{22} - \overline{U}_2) \\
&= q(1-q)\{p[[1 - \lambda(a+b)]\Delta r - (e + T_2S_2)] - C_2\}.
\end{aligned} \tag{13}$$

For  $F(q) = 0$ , the stable points are  $q_1^* = 0$ ,  $q_2^* = 1$ , and  $q_3^* = C_2/[1 - \lambda(a+b)]\Delta r - (e + T_2S_2)$ :

- (1) When  $p = C_2/[1 - \lambda(a+b)]\Delta r - (e + T_2S_2)$ , any value of  $q$  is an ESS
- (2) When  $p > C_2/[1 - \lambda(a+b)]\Delta r - (e + T_2S_2)$ , we get  $q_2^* = 1$ , an ESS
- (3) When  $p < C_2/[1 - \lambda(a+b)]\Delta r - (e + T_2S_2)$ , we get  $q_1^* = 0$ , an ESS

4.4.3. *Coevolution Strategy of MC and EF in the CVC Ecological Community.* By unifying the above various game results into a coordinate quadrant, we get comprehensive

game process of large companies and startups in CVC ecological community in Figure 9.

It can be seen from the figure that, in the regional QOPE, the evolutionary stability strategy of large companies and startups is (0,0); that is, neither party adopts the strategy of niche expansion in CVC activities. The evolutionary and stable strategies of both enterprises and startups are (1,1), which means that both large companies and startups are actively expanding their niche. However, the strategic actions of large enterprises and startups depend on the position of E. Through discussing the position of E, the factors that affect the strategic actions of large and startups can be identified.

$$\begin{aligned}
S_{QOPE} &= \frac{1}{2} \times 1 \times E_x + \frac{1}{2} \times 1 \times E_y = \frac{1}{2}(E_x + E_y) \\
&= \frac{1}{2} \left[ \frac{C_1}{\lambda(a+b)\Delta r - (e + T_1S_1)} + \frac{C_2}{[1 - \lambda(a+b)]\Delta r - (e + T_2S_2)} \right].
\end{aligned} \tag{14}$$

As long as the probability of more actions taken by large companies and startups on both sides of the game is within  $S_{QOPE}$ , both parties choose not to expand the niche. At this time, the evolution direction depends on  $S_{QOPE}$  and  $S_{QRPE}$ . When  $S_{QOPE} > S_{QRPE}$ , the evolution converges to point O. When  $S_{QOPE} < S_{QRPE}$ , the evolution converges to point B. When  $S_{QOPE} = S_{QRPE}$ , the probability of the two sides converging toward O or R is equal.

4.5. *Evolution Strategy of CVC Ecological Community.* Through the above analysis, 10 parameters that influence the direction of evolution are obtained. The common evolution path is discussed separately for each parameter by seeking partial derivatives:

- (1) Impact of  $C_1$  and  $C_2$  on  $S_{QOPE}$ :

$$\frac{\partial S}{\partial C_1} = \frac{1}{2} \times \frac{1}{\lambda(a+b)\Delta r - (e + T_1S_1)} > 0, \tag{15}$$

$$\frac{\partial S}{\partial C_2} = \frac{1}{2} \times \frac{1}{[1 - \lambda(a+b)]\Delta r - (e + T_2S_2)} > 0.$$

This shows that  $S_{QOPE}$  is a monotonically increasing function for  $C_1, C_2$ . For MC and EF, the higher the cost of forming a CVC ecological community is, the more it tends to maintain the original niche.

- (2) Impact of  $(a+b)$  on  $S_{QOPE}$ :

$$\begin{aligned}
\frac{\partial S}{\partial a} &= \frac{1}{2} \left[ \frac{C_2}{[[1 - \lambda(a+b)]\Delta r - (e + T_2S_2)]^2} \right. \\
&\quad \left. - \frac{C_1}{[\lambda(a+b)\Delta r - (e + T_1S_1)]^2} \right].
\end{aligned} \tag{16}$$

Since  $a$  is not monotonic for  $S_{QOPE}$ , a second-order derivative is performed:



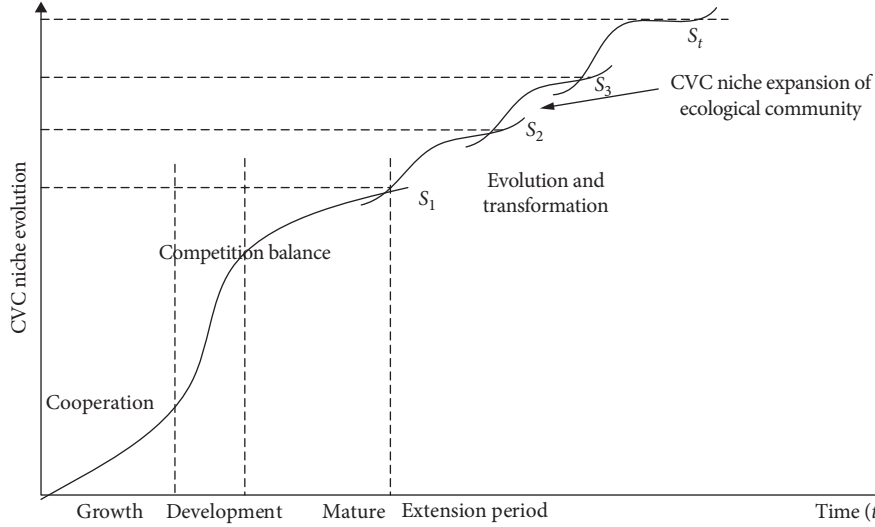


FIGURE 9: Comprehensive game process of large companies and startups in CVC ecological community after comprehensive situational factors.

$$\frac{\partial^2 S}{\partial^2 a} = \frac{C_2 \Delta r^2}{[[1 - \lambda(a+b)]\Delta r - (e + T_2 S_2)]^3} - \frac{C_1 \Delta r^3}{[\lambda(a+b)\Delta r - (e + T_1 S_1)]^3} > 0. \quad (17)$$

Therefore, there is a minimum value for  $S_{QOPE}$ .

When

$$\frac{\partial^2 S}{\partial^2 a} = \frac{C_2 \Delta r^2}{[[1 - \lambda(a+b)]\Delta r - (e + T_2 S_2)]^3} - \frac{C_1 \Delta r^3}{[\lambda(a+b)\Delta r - (e + T_1 S_1)]^3} = 0, \quad (18)$$

we obtain

$$\frac{C_2}{[[1 - \lambda(a+b)]\Delta r - (e + T_2 S_2)]^2} = \frac{C_1}{[\lambda(a+b)\Delta r - (e + T_1 S_1)]^2}. \quad (19)$$

At this time, the stable strategy of MC and EF is (1,1). Both of them implement niche expansion strategies. The parameter  $b$  is derived from the above  $a$ .

(3) Impact of  $T_1$ ,  $S_1$ ,  $T_2$ , and  $S_2$  on  $S_{QOPE}$ :

$$\begin{aligned} \frac{\partial S}{\partial T_1} &= \frac{1}{2} \frac{C_1}{[\lambda(a+b)\Delta r - (e + T_1 S_1)]^2} > 0, \\ \frac{\partial S}{\partial S_1} &= \frac{1}{2} \frac{C_1}{[\lambda(a+b)\Delta r - (e + T_1 S_1)]^2} > 0, \\ \frac{\partial S}{\partial T_2} &= \frac{1}{2} \frac{C_1}{[[1 - \lambda(a+b)]\Delta r - (e + T_2 S_2)]^2} > 0, \\ \frac{\partial S}{\partial S_2} &= \frac{1}{2} \frac{C_1}{[[1 - \lambda(a+b)]\Delta r - (e + T_2 S_2)]^2} > 0. \end{aligned} \quad (20)$$

$S_{QOPE}$  is a monotonically increasing function of  $T_1$ ,  $S_1$ ,  $T_2$ , and  $S_2$ , indicating that the larger the state and potential of the niche of large companies and startups in the CVC ecosystem (that is, the technology, talents, capital, and social resources of large companies and startups), the better the foundation, and the greater the real dominance and influence that large companies and startups have on the innovation environment. It is more likely that two sides tend to maintain their original niche.

(4) The impact of  $e$  on  $S_{QOPE}$ :

$$\frac{\partial S}{\partial e} = \frac{1}{2} \left[ \frac{C_2}{[[1 - \lambda(a+b)]\Delta r - (e + T_2 S_2)]^2} + \frac{C_1}{[\lambda(a+b)\Delta r - (e + T_1 S_1)]^2} \right] > 0. \quad (21)$$

It shows that  $S_{AECO}$  is a monotonically increasing function of  $e$ , which means that the greater the innovation income from maintaining the original niche is, the more likely the large enterprises and startups tend to maintain the original niche.

(5) The impact of  $\Delta r$  on  $S_{QOPE}$ :

$$\frac{\partial S}{\partial \Delta r} = \frac{1}{2} \left[ \frac{(1 - \lambda(a+b))C_2}{[[1 - \lambda(a+b)]\Delta r - (e + T_2 S_2)]^2} + \frac{\lambda(a+b)C_1}{[\lambda(a+b)\Delta r - (e + T_1 S_1)]^2} \right] < 0. \quad (22)$$

$S_{QOPE}$  is a monotonic reduction function for  $\Delta r$ ; that is, the larger the excess returns, the larger the result of the game between the large enterprise and the entrepreneurial enterprise. Both parties choose to expand the niche strategy.

(6) The impact of  $\lambda$  on  $S_{QOPE}$ :

$$\frac{\partial S}{\partial \lambda} = -\frac{1}{2} \left[ \frac{(1 - \lambda(a + b))C_2}{[[1 - \lambda(a + b)]\Delta r - (e + T_2S_2)]^2} + \frac{\lambda(a + b)C_1}{[\lambda(a + b)\Delta r - (e + T_1S_1)]^2} \right]. \quad (23)$$

$S_{QOPE}$  is a monotonic reduction function for  $\Delta r$ ; that is, the larger the excess returns, the larger the result of the game between the MC and the EF. Both parties choose to expand niche strategy.

## 5. Optimization Strategy for CVC Niche Evolution in Ecological Community

The evolution of the ecological niche of the CVC ecological community is subject to the combined effects of multiple mechanisms, and competition and symbiosis are affected by interest-driven and ecological balance. The benefit-driven mechanism is the inherent growth mechanism of the CVC ecological community under a certain environment of innovation and entrepreneurship. The source of motivation is the maximum pursuit of individual interests by large enterprises and entrepreneurial enterprises. Large enterprise populations and entrepreneurial enterprise populations integrate related resources through horizontal symbiosis, vertical symbiosis, or dynamic alliances to eliminate or reduce resource bottlenecks and promote the smooth development of large enterprises and individual entrepreneurial enterprises. Therefore, the CVC ecological community, as a core part of the innovation and entrepreneurship ecosystem, presents a nonlinear and exponential growth trend. As a self-protection mechanism of the ecological niche of the CVC ecological community, the ecological balance mechanism can ensure the self-sustainability and steady evolution of the CVC niche. In addition to the innovation, entrepreneurship, and ecological resources, the growth of the population is affected by material, energy, and information accumulation but is also balanced and stabilized by self-organization and self-regulation because of environmental factors.

The CVC ecological niche represents the status of the CVC ecological community with large companies and startups as the core to own and control human resources, technical resources, capital, and other social resources, as well as the adaptability to the innovative ecological environment and the comprehensive use of various resources. The healthy growth of CVC ecological communities requires specific analysis of internal and external environmental ecological factors, including limiting factors and interest factors. The limiting factor is the bottleneck that restricts the development of the CVC ecological community. Both the large enterprise population and the entrepreneurial enterprise population focus on easing the constraint of the limiting factor to enhance the evolution capacity of the entire CVC ecological community. At the same time, the CVC ecological community also has benefit factors as its

driving force for development. The enhanced interest factors of core populations can enable their respective niche evolution to achieve sustainable competitiveness and become an irreproducible and far-reaching advantage through the formation of core competitiveness. The large enterprise population and entrepreneurial enterprise population in the CVC ecological community are usually in different positions in the value chain. Therefore, the CVC ecological community should make full use of the symbiotic resources on the value chain, unblock the symbiotic channels, or form strategic alliances or virtual industries to achieve coevolution for the niche. The large enterprise population and the entrepreneurial enterprise population within the CVC ecological community implement corresponding niche separation strategies, niche expansion strategies, niche K-R strategies, and niche alliance strategies.

*5.1. CVC Niche Separation Strategy.* If the niche overlap between large enterprises and startups in the CVC ecological community indicates that there is a certain industry correlation between large companies and startups, the competition is fierce when the niche overlaps between the two. In order to obtain more living space, the two sides can implement niche separation strategy. Large enterprises and startups can find the most suitable position for their own development through CVC activities. The core of niche separation is to (a) use and integrate innovation and entrepreneurial resources, (b) choose a combination of ecological factors that are different from those of competitors, (c) have a stronger advantage at a certain position of talent, technology, capital, and social resource gradients, and (d) achieve heterogeneity and symmetry. Whether it is a mature large enterprise or a startup enterprise, it can choose two different niche separation strategies: specialization strategy and generalization strategy. The specialization strategy is for narrow-niche enterprises, which can provide a small range of products and services with characteristics to form an agglomeration strategy to concentrate the various resources of species in a limited ecological space, that is, to focus on a certain step in the value chain, so that the niche overlap with other species is reduced. The generalized strategy is similar to the no-difference strategy for relatively wide niche companies, and the risks are relatively small.

In the CVC ecological community, whether large companies and startups choose specialization or generalization strategies is related to the richness of resources they can use. Generalization strategies can provide more survival for large companies with strong competitiveness. However, early stage startups, which have limited resources while still implementing generalisation strategies, would easily fail or be acquired by large companies. . Therefore, for startup companies, specialization strategies are more conducive to maintaining their own niche and reducing the ecology overlap with large enterprises, to reduce competition. Startups in the CVC eco-community can also consider adopting a niche separation strategy from the time, space, and target market dimensions.

**5.2. CVC Niche Expansion Strategy.** The investment of mature large enterprises in startup enterprises through CVC activities is a niche expansion strategy. It uses open-ended innovation through venture capital to open up new resource spaces, acquire innovative technologies, and thus achieve the purpose of expanding the niche. They aim to alleviate the competitive pressure brought by overlapping niche. For large enterprises, the introduction of new and efficient ecological elements is their original intention when choosing CVC. It is also an effective means of exploring and developing suitable options for their own development so that they can acquire financial benefits. The introduction of new ecological elements can improve the utilization of various resources in the innovation and entrepreneurship ecosystem. Due to the obvious industry attributes of CVC and the high degree of innovation in the high-tech industry, the technology has a high rate of change; i.e., products and technologies will be updated faster. The uncertainty of independent research and development is high, so large enterprises need to expand their niche through CVC. For large enterprises and startups with weak innovation ability, weak interorganizational learning ability, and low absorptive capacity and involvement, especially for large companies, it is worthwhile to raise the basic niche or explore potential niche through CVC activities [17].

**5.3. CVC Niche K-R Strategy.** Ray proposed a niche selection strategy. The K strategy indicates that, by increasing competitiveness, improving resource utilization efficiency, and increasing the environmental capacity of species in a stable environment, a higher saturation density can be achieved. The R strategy means that, by breeding in large numbers, new offspring can be adapted to the unstable environment, thereby identifying its own niche [18].

The choice of K strategy will be limited by the environment. When the niche develops to a certain degree, it will reach the limit of affordability. In the CVC ecological community, the development of large enterprise populations also follows the KR strategy. The large enterprise population of the K strategy is characterized by a low reproduction rate and slow growth, but it can achieve strategic goals through CVC and continuously adapt to new environmental changes. It has unique core competitiveness in related industries. Its products and services are not easily copied and imitated. When large enterprises become the leading companies in an industry, environmental resources will be the biggest obstacle to their development [19]. The startups of R strategy are usually small in size, are not constrained by the environment, and have a fast reproduction capacity. When they grow to a certain degree, the growth rate also starts to decline.

The CVC ecological community uses the R strategy and the K strategy, respectively. The abscissa indicates the size of niche space occupied by the large enterprises and startups in the ecological community at time  $t$ , while the ordinate represents the size of the niche space occupied by species at time  $t + 1$ . The dotted line indicates that the size of the niche space has not changed ( $N_{t+1}/N_t = 1$ ). The part above the

dotted line indicates that the ecology of large enterprises or startups has become larger; that is,  $N_{t+1} > N_t$ . The curve below the dotted line indicates that the niche space has become smaller; that is  $N_{t+1} < N_t$  (Figure 10).

There are two intersections between the growth curve and the dotted line for K strategy. X is an unstable equilibrium point [20]. There may be two situations for large enterprises or startups at point X. First, when the CVC ecological community survives, there are large niche spaces that have not been occupied yet. Companies using K strategy to grow rapidly. Second, the overall space of the CVC ecological community has continued to shrink during the recession period, until the K strategy companies die out. When it is at a stable point, it indicates the maximum capacity of the ecosystem space and the maximum market capacity of technology, talent, capital, social resources, etc. When the K strategy prevents the individual from falling below stability, the scale is expanded to achieve stability. When the K strategy enterprise is above the stability point, its downsizing resumes [21]. The sustainable development of the CVC eco-community can be achieved if the large enterprise population and the entrepreneurial enterprise population in the CVC eco-community can adjust the niche countermeasures in a timely and accurate manner at different stages in the development process.

**5.4. CVC Niche Alliance Strategy.** In the CVC ecological community, large companies and startups can expand their niche by establishing a mutually beneficial symbiotic relationship. This strategy is called the niche alliance strategy. When the weaker startups and the stronger large companies can improve their respective competitiveness through symbiotic relationships, symbiotic parties in the CVC ecosystem can either choose their own appropriate niche or expand the boundaries of the entire ecological community from the perspective of enhancing the symbiotic system, promoting the metabolism of the entire CVC ecological community based on innovation. If the implementation of the niche alliance strategy requires the evaluation of their respective niche, the niche overlap and market crossover of both sides should be minimized to avoid parasitic or favorable symbiosis and form conflicts of interest [22]. Large enterprises and startups can seek intersections, that is, new requirements between the two sides. They can use the positive feedback effect of the innovation and entrepreneurship environment to create higher-level niche that is beneficial to both sides and can realize the coevolution of the niche. The main motivation of CVC activities is that large enterprises need to seek technological progress. Entrepreneurs in small and medium habitats in the industry can obtain more cooperation resources through the form of CVC, so that they can expand their survival through cooperative technological innovation, management innovation, and differentiated service innovation. In this case, they can expand the width of the niche and then penetrate or embed other industries. The formation and evolution of large enterprises and entrepreneurial enterprises in the CVC ecological

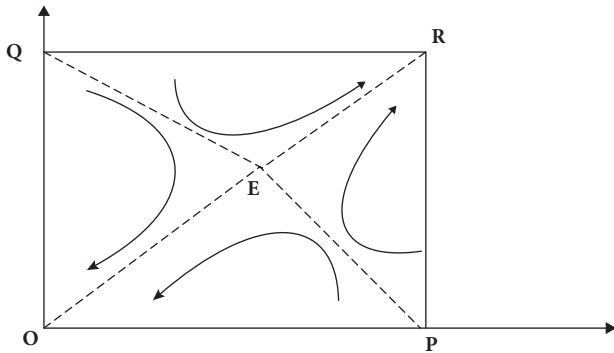


FIGURE 10: Coevolution of niche-cooperation relationships in CVC ecological communities.

community will also generate greater benefits in their own or other fields.

### Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Acknowledgments

This study was sponsored by the National Social Science Fund of China (Grant no. 18CGL015).

### References

- [1] T. Tykvová, "Legal framework quality and success of (different types of) venture capital investments," *Journal of Banking & Finance*, vol. 87, no. 2, pp. 333–350, 2018.
- [2] C. H. Dyer, G. P. Hammond, C. I. Jones, and R. C. McKenna, "Enabling technologies for industrial energy demand management," *Energy Policy*, vol. 36, no. 12, pp. 4434–4443, 2008.
- [3] D. P. Hannah and K. M. Eisenhardt, "Bottlenecks, cooperation, and competition in nascent ecosystems," *Strategic Management Journal*, vol. 40, no. 9, pp. 1333–1335, 2019.
- [4] L. Luo, P. K. Kannan, and B. T. Ratchford, "New product development under channel acceptance," *Marketing Science*, vol. 26, no. 2, pp. 149–163, 2007.
- [5] Y. Yao and H. Zhou, "The dynamic equilibrium and simulation of mobile internet platform innovation ecosystem," *Kybernetes*, vol. 45, no. 9, pp. 1406–1420, 2016.
- [6] D. Firoiu and A. G. Croitoru, "Tourism and tourism infrastructure from the perspective of technological changes," *Romanian Economic Business Review*, vol. 8, no. 8, pp. 93–103, 2013.
- [7] D. A. Levinthal, "Organizational adaptation and environmental selection-interrelated processes of change," *Organization Science*, vol. 2, no. 1, pp. 140–145, 1991.
- [8] M. Zaman Mir and A. Shiraz Rahaman, "The role of accounting in the enterprise bargaining process of an Australian university," *Accounting, Auditing & Accountability Journal*, vol. 16, no. 2, pp. 298–315, 2003.
- [9] V. Jog and B. J. Mcconomy, "Voluntary disclosure of management earnings forecasts in IPO prospectuses[J]," *Journal of Business Finance & Accounting*, vol. 30, no. 2, pp. 125–168, 2003.
- [10] Z. Jing and L. Xiong-Jian, "Business ecosystem strategies of mobile network operators in the 3G era: the case of China Mobile," *Telecommunications Policy*, vol. 35, no. 2, pp. 156–171, 2011.
- [11] T. Wei, Z. Zhu, Y. Li, and N. Yao, "The evolution of competition in innovation resource: a theoretical study based on Lotka–Volterra model," *Technology Analysis & Strategic Management*, vol. 30, no. 3, pp. 295–310, 2018.
- [12] S. K. Wilson, R. Fisher, M. S. Pratchett et al., "Habitat degradation and fishing effects on the size structure of coral reef fish communities," *Ecological Applications*, vol. 20, no. 2, pp. 442–451, 2010.
- [13] O. Honnay, B. Bossuyt, K. Verheyen, J. Butaye, H. Jacquemyn, and M. Hermy, "Ecological perspectives for the restoration of plant communities in European temperate forests," *Biodiversity and Conservation*, vol. 11, no. 2, pp. 213–242, 2002.
- [14] T. L. Vincent, M. V. Van, and B. S. Goh, "Ecological stability, evolutionary stability and the ESS maximum principle," *Evolutionary Ecology*, vol. 10, no. 6, pp. 567–591, 1996.
- [15] I. Loera, V. Sosa, and S. M. Ickert-Bond, "Diversification in North American arid lands: niche conservatism, divergence and expansion of habitat explain speciation in the genus *Ephedra*," *Molecular Phylogenetics and Evolution*, vol. 65, no. 2, pp. 437–450, 2012.
- [16] J. Li, Y. Huang, and X. Niu, "A branch population genetic algorithm for dual-resource constrained job shop scheduling problem," *Computers & Industrial Engineering*, vol. 102, no. 12, pp. 113–131, 2016.
- [17] B. Herault, "Reconciling niche and neutrality through the Emergent Group approach," *Perspectives in Plant Ecology, Evolution and Systematics*, vol. 9, no. 2, pp. 71–78, 2007.
- [18] S. K. Ray, "Effect of phosphorus on carbon activity, carbide precipitation, and coarsening in ferritic Fe–C–P alloys," *Metallurgical Transactions A*, vol. 22, no. 1, pp. 35–43, 1991.
- [19] B. Lavm and E. Nevo, "Genetic diversity in marine molluscs: a test of the niche-width variation hypothesis," *Marine Ecology*, vol. 2, no. 4, pp. 335–342, 1981.
- [20] C. Xu, "A novel recommendation method based on social network using matrix factorization technique," *Information Processing & Management*, vol. 54, no. 3, pp. 463–474, 2018.
- [21] Z. Chen, Y. Zhang, C. Wu, and B. Ran, "Understanding individualization driving states via latent dirichlet allocation model," *IEEE Intelligent Transportation Systems Magazine*, vol. 11, no. 2, pp. 41–53, 2019.
- [22] Z. Xie, R. Lin, J. Wang, W. Hu, and L. Miao, "Vicarious learning: how entrepreneurs enhance a Firm's international competitiveness through learning from interlocking director network partners," *Frontiers in Psychology*, vol. 11, p. 689, 2020.

## Review Article

# Multiobject Detection Algorithm Based on Adaptive Default Box Mechanism

Jinling Li <sup>1</sup>, Qingshan Hou <sup>2</sup>, and Jinsheng Xing <sup>2</sup>

<sup>1</sup>School of Economics and Management, Shanxi Normal University, Linfen 041004, China

<sup>2</sup>School of Mathematics and Computer Science, Shanxi Normal University, Linfen 041004, China

Correspondence should be addressed to Qingshan Hou; [hqs@stu.sxnu.edu.cn](mailto:hqs@stu.sxnu.edu.cn) and Jinsheng Xing; [ljli@sxnu.edu.cn](mailto:ljli@sxnu.edu.cn)

Received 15 July 2020; Revised 12 August 2020; Accepted 19 August 2020; Published 28 August 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Jinling Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Multiobject detection tasks in complex scenes have become an important research topic, which is the basis of other computer vision tasks. Considering the defects of the traditional single shot multibox detector (SSD) algorithm, such as poor small object detection effect, reliance on manual setting for default box generation, and insufficient semantic information of the low detection layer, the detection effect in complex scenes was not ideal. Aiming at the shortcomings of the SSD algorithm, an improved algorithm based on the adaptive default box mechanism (ADB) is proposed. The algorithm introduces the adaptive default box mechanism, which can improve the imbalance of positive and negative samples and avoid manually set default box super parameters. Experimental results show that, compared with the traditional SSD algorithm, the improved algorithm has a better detection effect and higher accuracy in complex scenes.

## 1. Introduction

With the continuous improvement of deep learning related theories, computer vision technologies [1–3] have achieved great success. As the basis of computer vision tasks, object detection [4–6] has been applied in many fields such as intelligent security [7], automatic driving [8], and intelligent medical treatment [9]; even some industrial applications are based on object detection algorithms [10–13]. In the past few years, in order to improve the real-time performance and accuracy of object detection in complex scenes, many scholars have conducted a lot of research on this, and the object detection algorithms based on deep learning have achieved remarkable achievements.

In 2014, the Region with CNN features (RCNN) algorithm [14] was published in Computer Vision and Pattern Recognition (CVPR) by Ross Girshick et al. The advent of this algorithm marked a new era in object detection technology. After that, Spatial Pyramid Pooling in Deep Convolutional Networks (SPPNet) algorithm [15] makes up for the shortcomings of the RCNN algorithm in repetitive convolution calculation and fixed output scale, but it still has

the defects of tedious training steps and slow process. In order to improve the real-time performance of the RCNN detection algorithm, Ross Girshick proposed the Fast RCNN object detection algorithm [16] in 2015. The processing mode of shared convolution makes the calculation amount of this algorithm drop sharply. In addition, the method of the region of interest (RoI) pooling is introduced to enable the network to process input images of any size. However, in this method, the problem of time loss caused by the selective search method [17] has not been solved. As a result, the Faster RCNN [18] algorithm was published in Neural Information Processing Systems (NIPS) in 2015. The highlight of this algorithm is to propose region proposal network (RPN) network structure, which combines region generation with convolution neural network based on the default box mechanism. It further improves the real-time performance of the Fast RCNN algorithm and becomes the most representative algorithm in the two-stage detection algorithm. Based on the Faster RCNN algorithm, Mask RCNN [19], fully convolutional network Region-based Fully Convolutional Networks (R-FCN) [20], Cascade RCNN [21], and other improved algorithms were proposed. Compared

with the two-stage structure of Faster RCNN and other algorithms, the You Only Look Once (YOLO) series of algorithms [22–25] and the SSD algorithm [26] adopt the single-stage structure to directly predict the location and category of the object. The real-time performance of the algorithm is greatly improved, but the accuracy of object detection in complex scenes is obviously insufficient. In order to improve the accuracy of the single-stage detection algorithm, Fu et al. [27] proposed a feature fusion method for multiscale prediction based on the SSD algorithm and used deconvolution operation to enhance the semantic information of shallow features. Jeong et al. [28] tried pooling fusion, deconvolution fusion, rainbow fusion, and other schemes and finally designed the Rainbow Single Shot Multibox Detector (RSSD) network model, which improved the object detection effect in complex scenes to some extent. Tsung et al. [29] considered that the reason for the lack of accuracy of the single-stage detection model was the imbalance of positive and negative samples and proposed a new classification loss function “Focal Loss” to improve the problem. Based on Focal Loss, the team of Tsung built a RetinaNet network, which effectively balanced the proportion of positive and negative samples to avoid the imbalance of samples.

This paper is based on the SSD algorithm, and we improve it from the following two aspects in view of the shortcomings of the SSD algorithm. On the one hand, in order to enhance the characterization capability of the low feature layers and the detection effect of small objects, the improved algorithm introduces feature layers fusion (FLF) and multireceptive field fusion (MRFF) mechanisms. On the other hand, through the adaptive default box mechanism, the steps of manually setting default box hyperparameters are avoided, the generation of negative sample box is reduced, and the problem of positive and negative sample imbalance is improved. Under the premise of real-time detection, the improved algorithm greatly improves the accuracy of small object detection in complex scenes.

## 2. Related Work

*2.1. SSD Algorithm.* The traditional SSD algorithm takes Visual Geometry Group Network (VGGNet) [30] as the backbone network and adds several additional convolution layers to participate in the detection of related objects. Firstly, sufficient data enhancements have been made through optical changes, geometric transformations, etc., which greatly enriched the relevant data sets. Secondly, the SSD algorithm expands 4 convolutional layers and performs object detection based on convolutional layers of different depths. Therefore, the output feature maps have different scales and receptive fields, and objects of different sizes can be detected. Thirdly, the SSD algorithm sets multiple default boxes of fixed size and ratio on the six feature maps. The algorithm sets a series of smaller default boxes on the shallow feature maps to detect small objects and sets several larger default boxes on the deep feature maps to detect large objects. Finally, the network model uses  $3 \times 3$  convolution kernels to extract features on the relevant feature maps to

complete objects classification and bounding boxes regression.

The SSD algorithm completes object detection through the single-stage network and has a better effect compared with the algorithm in the same period. Correspondingly, the SSD algorithm also has some shortcomings. On the one hand, due to the lack of semantic information in the shallow feature maps, the classification and regression effect of small objects is poor and the detection accuracy is insufficient. On the other hand, the default box parameters of each feature layer depend on the manual setting, so the generalization of the SSD is poor in different detection tasks.

*2.2. Design Criteria and Defects of Default Box in Traditional Detector.* A series of default boxes are generated by using the sliding window method in the relevant feature maps of the models, which is the mainstream method adopted by various object detection models at present. Firstly, the model defines several default boxes with specific scales and aspect ratios. Secondly, a large number of default boxes for object detection tasks are generated by sliding in the relevant output feature maps with a certain step size. The traditional default box generation method has the following disadvantages:

- (1) The traditional object detection models need to define a series of aspect ratios and scales for the default box. The selection of the default box aspect ratios and scales of the model will directly affect the detection effect of the models. In addition, for different data sets and detection methods, the parameters of the default box need to be adjusted according to the situation. If the selected default box parameters are not appropriate, the recall rate of the model will be too low, and the detection effect of object detection model will be poor
- (2) In the output feature maps of the relevant models, a large number of default boxes are distributed in the background area of the input image, which cannot play a good role in the detection of relevant objects
- (3) For the objects with a large difference in size and aspect ratio, a series of predefined default boxes may not be able to meet the detection requirements of the model
- (4) A large number of default boxes will directly lead to the degradation in the precision rate and real-time performance of the detection model

## 3. Improved Algorithm Design

*3.1. Fusion Mechanism in the Improved Algorithm.* In order to fuse the low-level output feature maps of different scales, on the one hand, the feature maps of Conv4\_3, FC7, and Conv6\_2 were dimensionally reduced to 256. The feature maps of FC7 and Conv6\_2 were adjusted to  $38 \times 38$  by bilinear interpolation method and concat operation is carried out for the relevant feature maps after processing. On the other hand, the dimension of the FC7 feature map is reduced to 512, the dimension of the Conv6\_2 output feature

map remains unchanged, and the dimension of Conv7\_2 is transformed to 512. The bilinear interpolation method is used to adjust the output feature maps of Conv6\_2 and Conv7\_2 to  $19 \times 19$ . Similarly, the concat operation is performed based on the processed feature maps.

The size of the human receptive field will change with the eccentricity of retinal imaging, and the size of the receptive field is proportional to the eccentricity. In order to further improve the detection efficiency of the traditional SSD object detection model, the multireceptive field fusion mechanism was added to the improved model by referring to the human visual perception mechanism.

Based on convolution kernels of different sizes and dilated convolution of different scales [31], the relevant mechanism fuses multiple scales of receptive fields to make the improved model have stronger feature expression ability. The fusion mechanism of multiple receptive fields in the improved model is shown in Figure 1.

The multireceptive field fusion mechanism consists of three branches. Firstly, convolution kernels of  $1 \times 1$ ,  $3 \times 3$ , and  $5 \times 5$  are used to simulate receptive fields of different sizes. Secondly, dilated convolution with rates of 1, 3, and 5 is used to simulate different degrees of eccentricity. In addition, for the  $3 \times 3$  and  $5 \times 5$  branches, the dimension of the feature map is reduced by using  $1 \times 1$  convolution kernel, so as to reduce the number of parameters. The feature map after dimension reduction is then sent into the convolution kernels of  $3 \times 3$  and  $5 \times 5$ . Finally, the fusion of 3 branches is completed by channel concat, and the number of feature channels is reduced by  $1 \times 1$  convolution kernel.

**3.2. Adaptive Default Box Mechanism.** The parameters' setting of the default box is the key part of the SSD object detection method. Similar to most mainstream object detection methods, the setting and generation of default boxes in the SSD algorithm also rely on the artificial unified setting. A series of preset default boxes are applied to the output features of relevant detection layers in the SSD algorithm. Since there are a large number of default boxes in the background area of the input image, and the aspect ratio of predefined default boxes may not be applicable to the objects to be detected in the relevant image, therefore, the detection efficiency of the model is greatly reduced by using this scheme.

The distribution of the objects in the input image is usually uneven, and the generation of default boxes is usually related to the content of the input image, the location, and the shape of the objects to be detected. Accordingly, the improved SSD object detection algorithm no longer uses the traditional default box generation strategy. The semantic information obtained by the algorithm is used to guide the generation of a series of appropriate size default boxes. The default box of an object is represented as  $(x, y, w, h)$ , where  $(x, y)$  represents the central coordinate position, and  $w$  and  $h$  represent the width and height, respectively. Assuming that  $A$  is an object to be detected on the input image  $G$ , the distribution of the corresponding default box can be represented by

$$A: p(x, y, w, h | G) = p(x, y | G)p(w, h | x, y, G). \quad (1)$$

According to equation (1), we can obtain two aspects of information. On the one hand, the object  $A$  to be detected may only appear in a partial area of the input image  $G$ . On the other hand, the distribution and scale of the corresponding default box are closely related to the location of object  $A$ . Therefore, the adaptive default box mechanism of the improved SSD model is shown in Figure 2.

The adaptive default box generation mechanism includes two parts: position prediction and shape prediction. Assuming that the input image is  $G$ , on the one hand, the position feature map is generated through the position prediction branch of the mechanism. The probability and position distribution of the objects to be detected in the input image can be obtained through the position feature map. On the other hand, according to the position prediction and shape prediction branches, the sizes and aspect ratios of the default boxes are predicted to generate the default boxes with different sizes and aspect ratios. Therefore, the default boxes in the improved SSD model are variable, and different contents can be obtained according to the features in different positions of the output feature maps. Considering that the shape of the default box is not fixed, by introducing the feature adaptive module, we carry out the feature adaptive adjustment for the improved model.

**3.3. Default Box Position and Shape Prediction.** In the process of position prediction, the improved SSD detection model first generates a series of location feature maps.

We assume that  $(i, j)$  is the coordinate of a point in the position feature map, and its probability value  $P$  corresponds to the coordinate  $Q$  in the input image, which can be expressed by

$$\begin{aligned} p(i, j | F_{\text{Conv} \in \{\text{Conv4}_3, \text{Conv6}, \text{Conv7}, \text{Conv8}_2, \text{Conv9}_2, \text{Conv10}_2, \text{Conv11}_2\}}) \\ \longleftrightarrow Q((i + 0.5)s, (j + 0.5)s), \end{aligned} \quad (2)$$

in which  $F_{\text{Conv}}$  represents the output feature map of a certain detection layer and  $s$  represents the step size of the output feature map. The  $1 \times 1$  convolution kernel is used to process the output feature map of the relevant detection layer, and the score map of the objects to be detected in the input image is obtained. The position prediction map of  $F_{\text{conv}}$  is further generated by the Sigmoid function. A certain probability threshold is set to identify the possible position of the object to be detected.

Based on the position prediction of the default box, the default bounding boxes of the objects to be detected are predicted by the shape prediction branch. According to the output feature map of the relevant detection layer, the shape prediction branch of the default box will predict the best default box shape at each location in the feature map. That is, by predicting the width and height of the default box, the maximum IoU value can be generated as far as possible with the nearest ground truth bounding box. Due to the fact that the range generated when directly predicting the width and

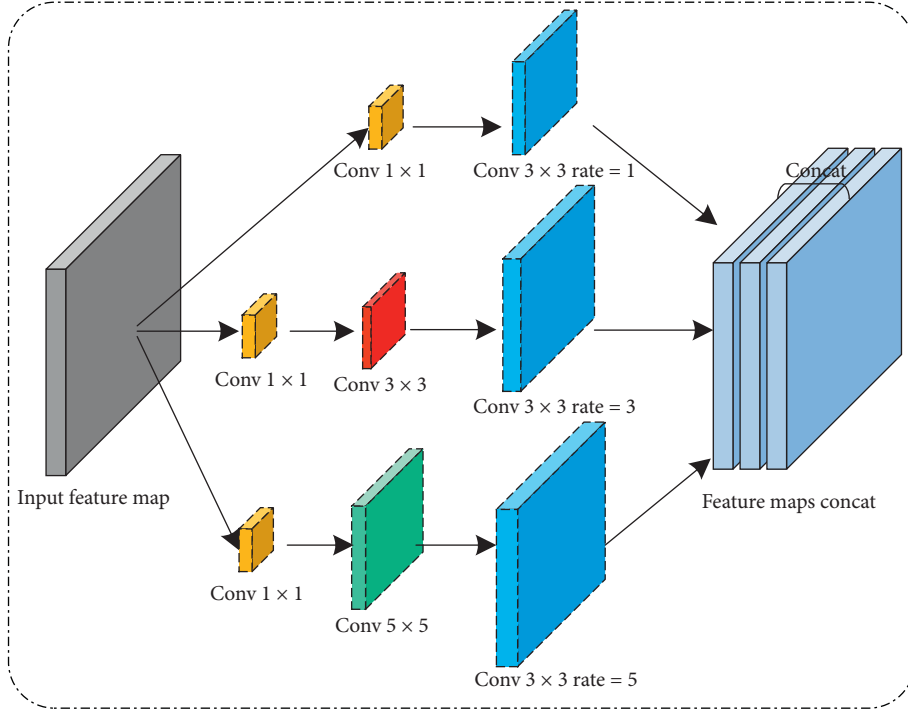


FIGURE 1: Multireceptive field fusion.

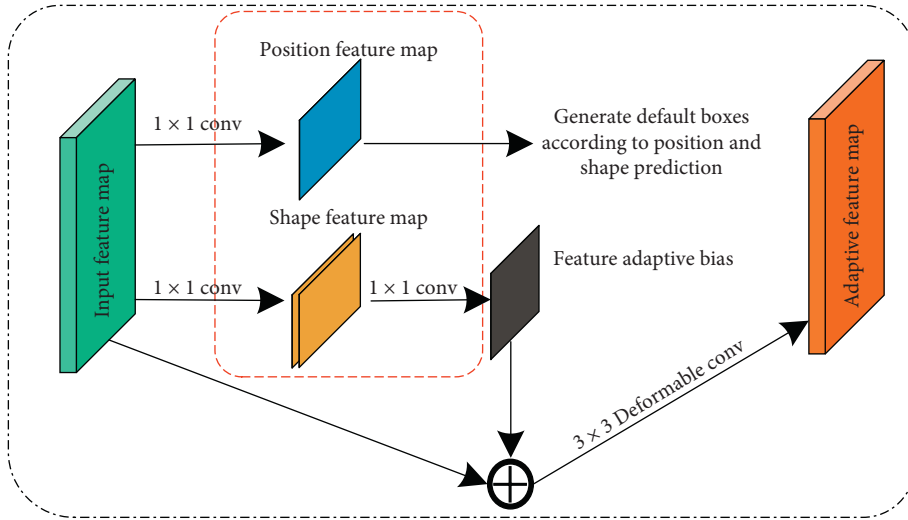


FIGURE 2: Adaptive default box mechanism for improved SSD model.

height of the bounding box is wide, and the prediction result is unstable, so it can be converted byt

$$w = \delta \cdot s \cdot e^{dw}, h = \delta \cdot s \cdot e^{dh}, \quad (3)$$

where  $s$  represents the step size and  $\delta$  represents the relative parameters controlling the default box size. Through equation (3), the output space can be mapped from  $[0, 1000]$  to  $[-1, 1]$ , so that the improved SSD object detection model can detect relevant objects more stably. The shape prediction branch uses the convolutional kernel of  $1 \times 1$  to predict the  $dw$  and  $dh$  values of the default box and completes the pixel-

level transformation of the relevant feature map through equation (3).

Compared with SSD, YOLO, RSSD, DSSD, and other object detection models, on the one hand, each position in the traditional models corresponds to a set of preset default bounding boxes. Each position in the feature maps of the improved model corresponds to only one prediction default box. The number of default boxes is greatly reduced, and the generated default boxes are more closely related to the objects to be detected. On the other hand, in the default box prediction scheme of the improved model, the aspect ratio of the default box does not need to be set manually. So, it also



has a better detection effect for the abnormal size objects existing in the input image.

**3.4. Feature Adaptive Module.** In most object detection networks such as SSD, RSSD, and DSSD, the sizes and aspect ratios of the default box are consistent at each position of the feature map. Therefore, the general convolution can be used to extract features in the output feature maps of the detection layers. Furthermore, the relevant features of each default bounding box are expressed. Compared with the existing SSD, RSSD, DSSD, and other object detectors, the default boxes with different shapes are automatically generated in the improved model. The output feature maps of the detection layers cannot predict the shape of its default boxes, but it is necessary to predict the categories and position offsets of these default boxes in the subsequent stages. That is to say, there is a mismatch between the default box and the features of the default box in the improved SSD model. In order to solve the above problem, the improved model introduces the relevant feature adaptive module and adjusts the relevant output feature maps according to equation (4) based on the default box shape of each position:

$$f'_i = N_T(f_i, w_i, h_i), \quad (4)$$

where  $f_i$  represents the feature at the  $i^{\text{th}}$  position in the output feature map and  $(w_i, h_i)$  represents the width and height of the default box corresponding to the  $i^{\text{th}}$  position. After the prediction of the default box, in order to realize the relevant position transformation and adapt to the shape of the default bounding box, a  $3 \times 3$  deformable convolution is applied to the output feature map to realize  $N_T$ . Different from the ordinary deformable convolution, the bias value in the feature adaptive module comes from the predicted default boundary box; that is,  $1 \times 1$  convolution kernel is used to act on the predicted default bounding box. From the perspective of specific functions, the feature adaptive module of the improved SSD model is similar to the RoI Pooling layer in the Faster RCNN algorithm. The structure of the improved SSD model is shown in Figure 3.

**3.5. Loss Function Setting.** Different from the traditional object detection models, the loss in the improved model includes not only the general classification loss  $L_{\text{cls}}$  and regression loss  $L_{\text{reg}}$  but also the position loss  $L_{\text{loc}}$  and shape loss  $L_{\text{shape}}$  during the default bounding box prediction. The final loss function can be expressed by equation (5), and the position loss and shape loss are balanced by the parameters  $\beta_1$  and  $\beta_2$ :

$$L_{\text{all}} = \beta_1 L_{\text{loc}} + \beta_2 L_{\text{shape}} + L_{\text{cls}} + L_{\text{reg}}, \quad (5)$$

where the classification loss  $L_{\text{cls}}$  adopts Cross Entropy (CE) loss [32] and the regression loss  $L_{\text{reg}}$  adopts smooth  $L_1$  loss.  $L_{\text{cls}}$  and  $L_{\text{reg}}$  can be expressed by

$$L_{\text{cls}} = \frac{1}{N_{\text{Pos}}} \left( \sum_{i \in \text{Pos}} \text{CE}(p_i, \hat{p}_i) + \sum_{i \in \text{Neg}} \text{CE}(p_i, \hat{p}_i) \right), \quad (6)$$

$$L_{\text{loc}} = \frac{1}{N_{\text{Pos}}} \sum_{i \in \text{Pos}} \sum_{m \in \{cx, cy, w, h\}} \text{smooth}_{L_1}(\tilde{l}_i^m - \hat{g}_i^m), \quad (7)$$

where  $p_i$  represents the probability that sample  $i$  is predicted to be of a certain class.  $\hat{p}_i$  indicates that the  $i^{\text{th}}$  sample belongs to a label of a certain category, and its value is 0 or 1.  $l$  and  $g$ , respectively, represent the deviation between the prediction box and the ground true box with the default box. When the default boundary box is generated, since the number of positive samples is smaller than that of negative samples, the focal loss is adopted to solve the problem of unbalanced positive and negative samples in position prediction. It can effectively reduce the loss of positive samples and the weight of negative samples in the training process. The loss can be expressed in equation (8). The value of the balance factor  $\alpha$  is set to 0.25 and the value of the regulation coefficient  $\gamma$  is set to 2:

$$L_{\text{Focalloss}} = \begin{cases} -\alpha(1-p)^\gamma \log p, & \hat{p} = 1, \\ -(1-\alpha)p^\gamma \log(1-p), & \hat{p} = 0. \end{cases} \quad (8)$$

When calculating the shape loss of the model,  $\text{IoU}_{\text{max}}$  is taken as a measure of the relevant loss. Based on the position feature map of each detection layer, several groups of different aspect ratios are sampled at each positive sample point position to complete the matching of IoU and determine the optimization object. Correlation matching can be expressed by equation (9). The shape loss of the improved model is shown in equation (10), where  $w_p$ ,  $h_p$ ,  $w_g$ , and  $h_g$ , respectively, represent the shapes of the prediction bounding box and the real bounding box:

$$\text{IoU}_{\text{max}} = \max_{w>0, h>0} \text{IoU}(\text{box}_{\text{wh}}, \text{box}_{\text{gt}}), \quad (9)$$

$$L_{\text{shape}} = \text{smooth } L_1 \left( 1 - \min \left( \frac{w_p}{w_g}, \frac{w_g}{w_p} \right) \right) + \text{smooth } L_1 \left( 1 - \min \left( \frac{h_p}{h_g}, \frac{h_g}{h_p} \right) \right). \quad (10)$$

## 4. Experimental Exploration

**4.1. Experimental Data Sets.** In view of the distribution of the objects in complex scenes, the ideal object detection model should have good generalization performance, which can not only effectively detect all kinds of different sizes of objects but also stably detect dense scenes. Based on Crowd Human [33], Pascal VOC 2012, and MS COCO data sets [34], relevant experiments verify the detection efficiency and robustness of the model on different data sets and comprehensively evaluate the improved model.

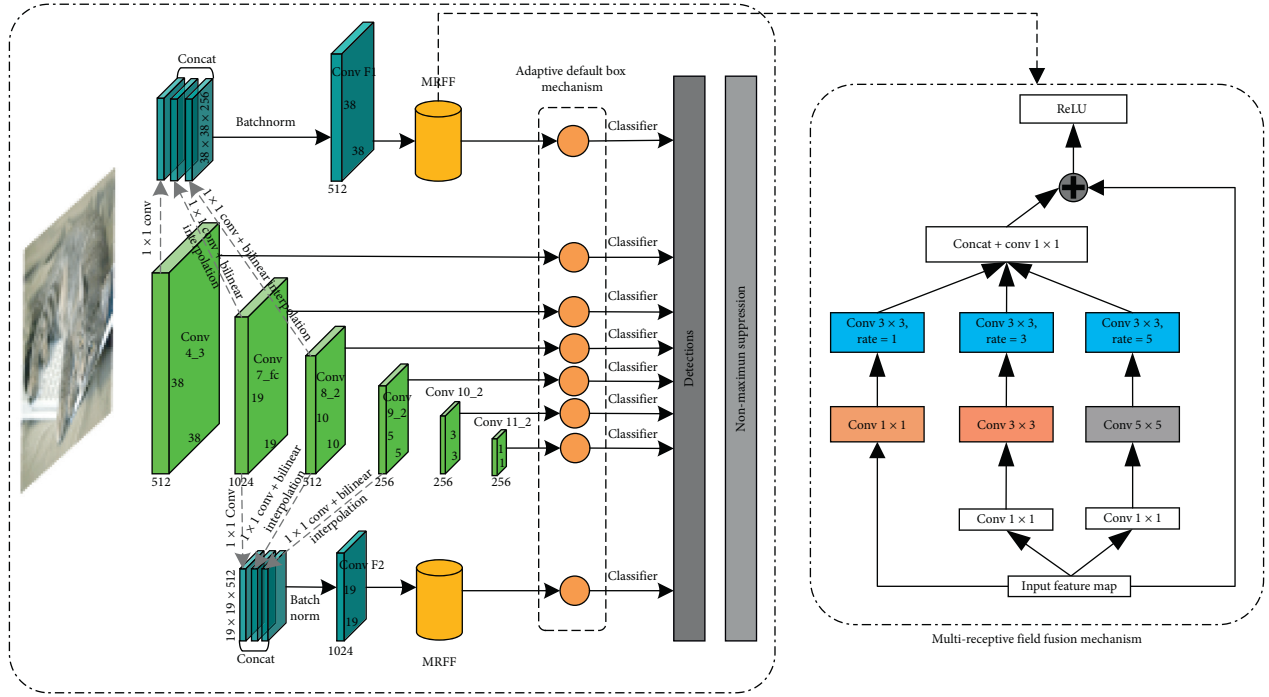


FIGURE 3: Improved SSD model structure.

The Crowd Human data set was released by MEGVII in 2018. The training set and validation set contain about 20,000 images. The data set contains more than 470,000 instances, and there are about 23 human instances in each image on average. At the same time, complex occlusion phenomenon exists in the images.

PASCAL VOC is a classic public data set, which mainly includes VOC 2007 and VOC 2012. This data set provides a complete set of standards for image classification and object detection. The VOC 2007 data set is extended to the VOC 2007 data set, which contains 20 kinds of objects and 11530 images. The related image annotations are complete and of high quality.

MS COCO data set is funded and annotated by Microsoft. It involves multiple computer vision tasks such as object detection, object segmentation, and semantic understanding. It contains about 300,000 data images, more than 2 million instances, and 91 kinds of objects. Compared with other public data sets, the COCO data set has more small objects, more complex object types, and detection scenarios. It can comprehensively evaluate model performance.

**4.2. Data Preprocessing and Model Evaluation Indexes.** In order to fully train the improved model, enhance the generalization of the model, and improve the detection effect of small objects and occlusion objects, the corresponding data preprocessing strategy is formulated. Generally, the object whose number of pixels is less than 1024 in the segmentation mask of the image object region is defined as a small object. Objects with more than 1024 pixels and less than 9216 pixels in the segmentation mask of the image object region are

defined as medium-size objects. It mainly includes two aspects: optical transformation and geometric transformation. Optical transformation mainly includes the adjustment of brightness, contrast, hue, saturation, and channel. The geometric transformation utilizes operations such as random cropping, random expansion, and scaling to achieve image size changes.

The performance of the improved model is measured by average precision (AP), average recall (AR), and frame per second (FPS). As the common evaluation indexes, the AP value reflects the precision and recall rate of the test results. The larger the value is, the better the detection precision of the model will be, and the AR value reflects the recall rate and positioning accuracy of the model. In addition to the detection precision, the FPS value is used to measure the detection speed of the improved algorithm, that is, the number of images the model can process per second.

**4.3. Model Parameters' Setting and Training.** The relevant models are trained and tested on Crowd Human, PASCAL VOC 2012, and MS COCO data sets, respectively, to verify the generalization performance of the improved model on different data sets. In the multitask loss function,  $\beta_1 = 1$  and  $\beta_2 = 0.1$  are set to balance the position loss and shape loss of the default box. The training of the model is based on the stochastic gradient descent algorithm and the "warm-up" strategy is adopted. During the initial five epochs, the learning rate of the model is increased from  $10^{-4}$  to  $4 * 10^{-3}$ . After the "warm-up" phase, the learning rate is changed to  $10^{-4}$  again, and the learning rate is set as  $10^{-5}$  and  $10^{-6}$ , respectively, at the 8th epoch and the 11th epoch. The momentum value during the training process is 0.9, and the

TABLE 1: Influence of relevant mechanisms on detection results.

| Data set        | Algorithms           | Conv 4_3 |        |     | Conv 7_fc |        |     | Conv F1 |        |     | Conv F2 |        |     |
|-----------------|----------------------|----------|--------|-----|-----------|--------|-----|---------|--------|-----|---------|--------|-----|
|                 |                      | AP (%)   | AR (%) | FPS | AP (%)    | AR (%) | FPS | AP (%)  | AR (%) | FPS | AP (%)  | AR (%) | FPS |
| Crowd human     | SSD                  | 84.6     | 72.3   | 59  | 86.4      | 73.5   | 59  | —       | —      | —   | —       | —      | —   |
|                 | SSD + FLF            | —        | —      | —   | —         | —      | —   | 86.3    | 74.1   | 50  | 87.7    | 75.4   | 50  |
|                 | SSD + ADB            | 92.1     | 81.6   | 50  | 93.6      | 82.4   | 50  | —       | —      | —   | —       | —      | —   |
|                 | SSD + FLF + MR       | —        | —      | —   | —         | —      | —   | 87.5    | 75.2   | 50  | 88.6    | 76.7   | 50  |
|                 | SSD + FLF + MR + ADB | —        | —      | —   | —         | —      | —   | 94.8    | 82.5   | 50  | 95.3    | 83.4   | 50  |
| PASCAL VOC 2012 | SSD                  | 65.3     | 55.2   | 56  | 67.2      | 56.3   | 56  | —       | —      | —   | —       | —      | —   |
|                 | SSD + FLF            | —        | —      | —   | —         | —      | —   | 67.1    | 57.3   | 47  | 68.4    | 58.3   | 47  |
|                 | SSD + ADB            | 72.6     | 61.4   | 47  | 74.8      | 62.7   | 47  | —       | —      | —   | —       | —      | —   |
|                 | SSD + FLF + MR       | —        | —      | —   | —         | —      | —   | 68.3    | 58.6   | 47  | 69.6    | 59.7   | 47  |
|                 | SSD + FLF + MR + ADB | —        | —      | —   | —         | —      | —   | 74.5    | 62.4   | 47  | 75.3    | 63.4   | 47  |
| MS COCO         | SSD                  | 38.1     | 30.2   | 54  | 39.4      | 31.4   | 54  | —       | —      | —   | —       | —      | —   |
|                 | SSD + FLF            | —        | —      | —   | —         | —      | —   | 40.3    | 32.6   | 44  | 41.6    | 33.8   | 44  |
|                 | SSD + ADB            | 45.3     | 36.1   | 44  | 46.8      | 37.3   | 44  | —       | —      | —   | —       | —      | —   |
|                 | SSD + FLF + MR       | —        | —      | —   | —         | —      | —   | 41.2    | 34.1   | 44  | 43.2    | 35.6   | 44  |
|                 | SSD + FLF + MR + ADB | —        | —      | —   | —         | —      | —   | 47.6    | 39.2   | 44  | 48.3    | 40.1   | 44  |

weight attenuation value is 0.005. All experiments are conducted based on GTX 1080 Ti and Titan X GPU environments.

#### 4.4. Influence of Relevant Mechanisms on the Detection Effect.

The improved SSD detection model is based on multiple feature maps for object detection. The deep feature maps with a large receptor field are responsible for the detection of large-scale objects, while the low-level feature maps with a small receptor field are responsible for the detection of small objects. By introducing the corresponding fusion mechanisms, the semantic information of low feature layers can be enriched. Accordingly, Conv 4\_3, Conv7\_fc, Conv F1, and Conv F2 are used for the detection of small objects, while the rest of the detection layers are used for the detection of larger objects. In addition, the ADB mechanism is added to the improved SSD model to improve the positioning precision of the model, avoid manually setting default box hyperparameters, and improve the imbalance of positive and negative samples. Based on different data sets, the experiment explored the influence of relevant mechanisms on the detection effect, and the experimental results are shown in Table 1.

Based on different data sets, Table 1 explores the influence of relevant mechanisms on the detection results of Conv 4\_3, Conv7\_fc, Conv F1, and Conv F2 layers. For Conv4\_3 layer, ADB mechanism was added to the improved algorithm, and the AP values of Crowd Human, PASCAL VOC 2012, and MS COCO data sets reached 92.1%, 72.6%, and 45.3%, respectively. AR value can be up to 81.6%, 61.4%, and 36.1%; compared with the traditional SSD algorithm, the average precision value and the average recall rate are greatly improved. In addition, it can be seen that the detection effect of Conv7\_fc has also been significantly improved.

In order to enhance the detection effect of small objects in dense scenes, additional small object detection layers Conv F1 and Conv F2 are added in the improved algorithm. The relevant detection layers use FLF, MRFF, and ADB generation mechanism to strengthen the semantic information

of the low detection layers. In the case of applying FLF and MR, the average detection precision of the Conv F1 detection layer on the three data sets can reach 87.5%, 68.3%, and 41.2%, respectively. Compared with the detection effect of the Conv4\_3 layer in the traditional SSD algorithm, the algorithm precision is improved. After the introduction of the ADB mechanism, the average detection precision and average recall rate of the algorithm are greatly improved. The experiment shows that the improved network has stronger characterization ability, better detection effect, and higher object positioning precision. Figure 4 shows the influence of relevant mechanisms on the detection effect. With the introduction of ADB, FLF, and MR, the low detection layers of the improved algorithm can extract richer feature information and detect more small objects compared with the original algorithm.

4.5. Comparison of Relevant Models. Based on the PASCAL VOC2012 test set, we compared the detection effects of Faster RCNN, YOLO V2, SSD, DSSD, RSSD, and our SSD algorithms. The training of the algorithm involved VOC 2012 and MS COCO training sets. The basic network included VGGNet, ResNet-101 [35], and Darknet-19. Taking FPS, mAP, and mAR [36] as evaluation criteria, the experimental comparison results of the six models are shown in Table 2.

By analyzing the experimental data in Table 2, our SSD300 has improved its average precision and average recall rate compared with Faster RCNN, YOLOv2, SSD300, DSSD321, and RSSD300 algorithms. The detection precision of our SSD300<sub>-s0</sub> model can reach 73.2% without pre-training, which is 3.6% higher than that of SSD300<sub>-s0</sub>. When the model training is combined with the MS COCO data set, the detection accuracy of our SSD300<sub>+coco</sub> reaches 83.4%, which is 2.2% higher than SSD300<sub>+coco</sub>. In addition, the average recall rate of our SSD300<sub>+coco</sub> is 74.1%, which is about 2.5% higher than SSD300<sub>+coco</sub>. This verifies the effectiveness of ADB and other relevant mechanisms, improves the imbalance of positive and negative samples in traditional SSD algorithms, and improves the detection effect of objects in dense scenes.



FIGURE 4: Influence of the relevant mechanism on small object detection layers. (a) Conv4\_3: SSD, (b) Conv4\_3: SSD + ADB, and (c) Conv F1: SSD + FLF + MR + ADB.

TABLE 2: Results based on the PASCAL VOC 2012 test set.

| Algorithm                    | Training set    | Pretraining | Basic network | FPS | GPU     | Input image            | mAP (%) | mAR (%) |
|------------------------------|-----------------|-------------|---------------|-----|---------|------------------------|---------|---------|
| Faster RCNN [18]             | VOC 2012        | Yes         | VGGNet        | 9   | Titan X | $\sim 600 \times 1000$ | 73.2    | 61.3    |
| Faster RCNN [18]             | VOC 2012        | Yes         | ResNet-101    | 3   | Titan X | $\sim 600 \times 1000$ | 76.4    | 63.5    |
| YOLO v2 [23]                 | VOC 2012        | Yes         | Darknet-19    | 81  | Titan X | $352 \times 352$       | 73.7    | 61.7    |
| SSD300 <sub>-so</sub> [26]   | VOC 2012        | No          | VGGNet        | 56  | 1080 Ti | $300 \times 300$       | 69.6    | 58.2    |
| SSD300 [26]                  | VOC 2012        | Yes         | VGGNet        | 56  | 1080 Ti | $300 \times 300$       | 77.2    | 68.3    |
| SSD300 <sub>+coco</sub> [26] | VOC 2012 + COCO | Yes         | VGGNet        | 56  | 1080 Ti | $300 \times 300$       | 81.2    | 71.6    |
| SSD512 [22]                  | VOC 2012        | Yes         | VGGNet        | 38  | 1080 Ti | $512 \times 512$       | 78.5    | 69.4    |
| SSD512 <sub>+coco</sub> [26] | VOC 2012 + COCO | Yes         | VGGNet        | 38  | 1080 Ti | $512 \times 512$       | 83.2    | 72.5    |
| DSSD321 [27]                 | VOC 2012        | Yes         | ResNet-101    | 12  | 1080 Ti | $321 \times 321$       | 78.6    | 69.7    |
| DSSD513 [27]                 | VOC 2012        | Yes         | ResNet-101    | 7   | 1080 Ti | $513 \times 513$       | 81.5    | 72.4    |
| RSSD300 [28]                 | VOC 2012        | Yes         | VGGNet        | 41  | 1080 Ti | $300 \times 300$       | 78.5    | 68.9    |
| RSSD512 [28]                 | VOC 2012        | Yes         | VGGNet        | 23  | 1080 Ti | $512 \times 512$       | 80.8    | 71.2    |
| Our SSD300 <sub>-so</sub>    | VOC 2012        | No          | VGGNet        | 47  | 1080 Ti | $300 \times 300$       | 73.2    | 62.4    |
| Our SSD300                   | VOC 2012        | Yes         | VGGNet        | 47  | 1080 Ti | $300 \times 300$       | 79.8    | 71.2    |
| Our SSD300 <sub>+coco</sub>  | VOC 2012 + COCO | Yes         | VGGNet        | 47  | 1080 Ti | $300 \times 300$       | 83.4    | 74.1    |
| Our SSD512                   | VOC 2012        | Yes         | VGGNet        | 30  | 1080 Ti | $512 \times 512$       | 82.4    | 73.6    |
| Our SSD512 <sub>+coco</sub>  | VOC 2012 + COCO | Yes         | VGGNet        | 30  | 1080 Ti | $512 \times 512$       | 85.6    | 75.9    |

Due to the introduction of ADB and other modules, some detection time is lost. Compared with the original SSD300, the real-time detection performance of the improved algorithm is reduced, but it is enough to meet the requirements of real-time detection. Compared with RSSD300, DSSD321, and other improved algorithms, our SSD300 not only improves the average detection precision

and average recall rate but also has obvious advantages in real-time detection.

In order to further verify the generalization of the improved algorithm on different data sets, a series of AP values and AR values of the relevant models were fully explored based on the MS COCO data set. The relevant experimental results are shown in Tables 3 and 4.

TABLE 3: AP results based on the MS COCO test set.

| Algorithm        | Training set     | Basic network | Pretraining | AP <sub>IoU</sub> (%)  |                   |                    | AP <sub>Area</sub> (%) |                 |                 |
|------------------|------------------|---------------|-------------|------------------------|-------------------|--------------------|------------------------|-----------------|-----------------|
|                  |                  |               |             | AP <sub>0.5-0.95</sub> | AP <sub>0.5</sub> | AP <sub>0.75</sub> | AP <sub>S</sub>        | AP <sub>M</sub> | AP <sub>L</sub> |
| Faster RCNN [18] | COCO trainval    | VGGNet        | Yes         | 21.9                   | 42.7              | —                  | —                      | —               | —               |
| YOLO v2 [23]     | COCO trainval35k | Darknet-19    | Yes         | 21.6                   | 44                | 19.2               | 5                      | 22.4            | 35.5            |
| SSD300 [26]      | COCO trainval35k | VGGNet        | Yes         | 25.1                   | 43.1              | 25.8               | 6.6                    | 25.9            | 41.4            |
| SSD512 [26]      | COCO trainval35k | VGGNet        | Yes         | 28.8                   | 48.5              | 30.3               | 10.9                   | 31.8            | 43.5            |
| DSSD321 [27]     | COCO trainval35k | ResNet-101    | Yes         | 28                     | 46.1              | 29.2               | 7.4                    | 28.1            | 47.6            |
| DSSD513 [27]     | COCO trainval35k | ResNet-101    | Yes         | 33.2                   | 53.3              | 35.2               | 13                     | 35.4            | 51.1            |
| RSSD300 [28]     | COCO trainval35k | VGGNet        | Yes         | 27.3                   | 45.8              | 28.9               | 7.2                    | 27.8            | 46.8            |
| RSSD512 [28]     | COCO trainval35k | VGGNet        | Yes         | 32.8                   | 52.7              | 34.7               | 12.4                   | 34.6            | 50.7            |
| Our SSD300       | COCO trainval35k | VGGNet        | Yes         | 29.4                   | 47.3              | 30.8               | 9.2                    | 30.3            | 49.3            |
| Our SSD512       | COCO trainval35k | VGGNet        | Yes         | 35.3                   | 54.8              | 36.4               | 14.3                   | 37.2            | 53.2            |

TABLE 4: AR results based on the MS COCO test set.

| Algorithm    | Training set     | Basic network | Pretraining | AR <sub>Detections</sub> (%) |                 |                  | AR <sub>Area</sub> (%) |                 |                 |
|--------------|------------------|---------------|-------------|------------------------------|-----------------|------------------|------------------------|-----------------|-----------------|
|              |                  |               |             | AR <sub>5</sub>              | AR <sub>8</sub> | AR <sub>12</sub> | AR <sub>S</sub>        | AR <sub>M</sub> | AR <sub>L</sub> |
| YOLO v2 [23] | COCO trainval35k | Darknet-19    | Yes         | 25.7                         | 30.6            | 32.3             | 9.8                    | 36.5            | 54.4            |
| SSD300 [26]  | COCO trainval35k | VGGNet        | Yes         | 28.7                         | 33.1            | 36.8             | 11.2                   | 40.4            | 58.4            |
| SSD512 [26]  | COCO trainval35k | VGGNet        | Yes         | 31.1                         | 37.5            | 41.4             | 16.5                   | 46.6            | 60.8            |
| DSSD321 [27] | COCO trainval35k | ResNet-101    | Yes         | 30.5                         | 35.1            | 38.9             | 12.7                   | 42              | 62.6            |
| DSSD513 [27] | COCO trainval35k | ResNet-101    | Yes         | 33.9                         | 42.5            | 45.8             | 21.8                   | 49.1            | 66.4            |
| RSSD300 [28] | COCO trainval35k | VGGNet        | Yes         | 30.2                         | 34.6            | 38.2             | 12.1                   | 41.2            | 61.8            |
| RSSD512 [28] | COCO trainval35k | VGGNet        | Yes         | 33.4                         | 41.8            | 44.8             | 20.9                   | 48.4            | 66.1            |
| Our SSD300   | COCO trainval35k | VGGNet        | Yes         | 32.4                         | 36.8            | 40.2             | 14.2                   | 43.8            | 64.3            |
| Our SSD512   | COCO trainval35k | VGGNet        | Yes         | 36.2                         | 44.2            | 47.2             | 23.6                   | 51.3            | 68.3            |

According to the experimental data in Tables 3 and 4, compared with Faster RCNN, YOLO V2, SSD, DSSD, and RSSD algorithms, our SSD still has good detection performance on MS COCO data set. In the detection of small objects, AP<sub>S</sub> and AR<sub>S</sub> of our SSD512 can reach 14.3% and 23.6%, respectively. Compared with the original SSD algorithm, the average detection precision and average recall rate of small objects have been improved by about 3.4% and 7.1%, respectively. In addition, the other evaluation indicators also have different degrees of improvement. The improved algorithm achieves ideal detection results on both MS COCO and PASCAL VOC data sets. On the one hand, the improved SSD algorithm has good generalization. On the other hand, it also directly shows the effectiveness of the algorithm improvement.

## 5. Conclusion

In view of the defects of the traditional SSD detection algorithm, such as the poor detection effect of small objects and the default box generation depending on manual settings, this paper proposes an improved multiobject detection algorithm, which effectively improves the object detection effect in complex scenes. The improved algorithm mainly involves the following contributions: on the one hand, the introduction of feature fusion and multireceptive field fusion mechanism enhances the characterization ability of the low feature layers and improves the detection effect of small objects. On the other hand, through the adaptive default box

mechanism, the steps of setting default box hyperparameters are avoided, the generation of negative sample box is reduced, and the imbalance of positive and negative samples is improved. Under the requirement of real-time detection, the improved algorithm greatly improves the average precision and recall rate of object detection in complex scenes.

## Data Availability

The image data supporting this algorithm research is from previously reported studies and datasets. These prior studies and datasets are cited at relevant places within the text as references [22–24, 29, 30].

## Conflicts of Interest

The authors declare that they have no conflicts of interest in this work.

## Authors' Contributions

All authors contributed equally to this manuscript.

## Acknowledgments

This work was supported in part by the Soft Science Foundation of Shanxi Province under Grant 2011041037-02 and in part by the Scholarship Council of Shanxi Province under Grant 2011-8.

## References

- [1] S. Ghosh, N. Das, I. Das, and U. Maulik, "Understanding deep learning techniques for image segmentation," *ACM Computing Surveys*, vol. 52, no. 4, pp. 1–35, 2019.
- [2] S. Asgari Taghanaki, K. Abhishek, J. P. Cohen, J. Cohen-Adad, and G. Hamarneh, "Deep semantic segmentation of natural and medical images: a review," *Artificial Intelligence Review*, 2020.
- [3] C. Chen, C. Qin, H. Qiu et al., "Deep learning for cardiac image segmentation: a review," *Frontiers in Cardiovascular Medicine*, vol. 7, 2020.
- [4] Z. Zou, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: a survey," 2019, <https://arxiv.org/abs/1905.05055>.
- [5] J. Li, Q. Hou, J. Xing, and J. Ju, "SSD object detection model based on multi-frequency feature theory," *IEEE Access*, vol. 8, no. 1, pp. 82294–82305, 2020.
- [6] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: a review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3212–3232, 2019.
- [7] Z. Lv, W. Kong, X. Zhang, D. Jiang, H. Lv, and X. Lu, "Intelligent security planning for regional distributed energy internet," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 5, pp. 3540–3547, 2020.
- [8] K. Chen, S. Tripathi, Y. Hwang, and T. Nguyen, "Moving objects detection using classifying object proposals for driver assistance system," in *Proceedings of the International SoC Design Conference (ISOCC)*, pp. 177–178, Jeju, Republic of Korea, October 2016.
- [9] S. Hussein, P. Kandel, C. W. Bolan, M. B. Wallace, and U. Bagci, "Lung and pancreatic tumor characterization in the deep learning era: novel supervised and unsupervised learning approaches," *IEEE Transactions on Medical Imaging*, vol. 38, no. 8, pp. 1777–1787, 2019.
- [10] H. Liang, J. Zou, K. Zuo, and M. J. Khan, "An improved genetic algorithm optimization fuzzy controller applied to the wellhead back pressure control system," *Mechanical Systems and Signal Processing*, vol. 142, Article ID 106708, 2020.
- [11] F. Hu and G. Wu, "Distributed error correction of EKF algorithm in multi-sensor fusion localization model," *IEEE Access*, vol. 8, pp. 93211–93218, 2020.
- [12] D. Jiang, G. Li, Y. Sun, J. Kong, and B. Tao, "Gesture recognition based on skeletonization algorithm and CNN with ASL database," *Multimedia Tools and Applications*, vol. 78, no. 21, pp. 29953–29970, 2019.
- [13] H. Liang, J. Zou, Z. Li, M. J. Khan, and Y. Lu, "Dynamic evaluation of drilling leakage risk based on fuzzy theory and PSO-SVR algorithm," *Future Generation Computer Systems*, vol. 95, pp. 454–466, 2019.
- [14] R. Girshick, J. Donahue, T. Darrell, and T. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, IEEE, Columbus, OH, USA, pp. 580–587, June 2014.
- [15] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial Pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.
- [16] R. Girshick, "Fast R-CNN," in *Proceedings of the International Conference on Computer Vision*, IEEE, Santiago, Chile, pp. 1440–1448, December 2015.
- [17] J. R. R. Uijlings, K. E. A. Van De Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *International Journal of Computer Vision*, vol. 104, no. 2, pp. 154–171, 2013.
- [18] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," in *Proceedings of the Advances in Neural Information Processing Systems*, Curran Associates Inc, Montreal, Canada, pp. 91–99, December 2015.
- [19] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proceedings of the International Conference on Computer Vision*, IEEE, Venice, Italy, pp. 2961–2969, October 2017.
- [20] J. Dai and Y. Li, "R-FCN: object detection via region-based fully convolutional networks," in *Proceedings of the Advances in Neural Information Processing Systems*, vol. 29, pp. 379–387, NIPS, Barcelona, Spain, December 2016.
- [21] Z. Cai and N. Vasconcelos, "Cascade R-CNN: delving into high quality object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Salt Lake City, UT, USA, pp. 6154–6162, June 2018.
- [22] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of the Computer Vision and Pattern Recognition*, IEEE, Las Vegas, NV, USA, pp. 779–788, June 2016.
- [23] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the Computer Vision and Pattern Recognition*, IEEE, Honolulu, HI, USA, pp. 7263–7271, July 2017.
- [24] J. Redmon and A. Farhadi, "YOLOv3: an incremental improvement," 2018, <https://arxiv.org/abs/1804.02767>.
- [25] A. Bochkovskiy, C.-Y. Wang, and H. M. Liao, "YOLOv4: optimal speed and accuracy of object detection," 2020, <https://arxiv.org/abs/2004.10934>.
- [26] W. Liu, D. Anguelov, D. Erhan et al., "SSD: single shot multibox detector," in *Proceedings of the European Conference on Computer Vision*, Springer, Cham, Switzerland, pp. 21–37, October 2016.
- [27] C. Y. Fu, W. Liu, A. Ranga, A. Tyagi, and A. C. Berg, "DSSD: deconvolutional single shot detector," 2017, <https://arxiv.org/abs/1701.06659>.
- [28] J. Jeong, H. Park, and N. Kwak, "Enhancement of SSD by concatenating feature maps for object detection," in *Proceedings of the British Machine Vision Conference*, BMVA Press, London, UK, May 2017.
- [29] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the International Conference on Computer Vision*, IEEE, Venice, Italy, pp. 2980–2988, October 2017.
- [30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proceedings of the International Conference on Learning Representations 2015*, San Diego, CA, USA, May 2015.
- [31] L.-C. Chen, P. George, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, <https://arxiv.org/abs/1706.05587>.
- [32] G. E. Nasr, E. A. Badr, and C. Joun, "Cross entropy error function in neural networks: forecasting gasoline demand," in *Proceedings of the Fifteenth International Florida Artificial Intelligence Research Society Conference*, AAAI Press, Pensacola Beach, FL, USA, May 2002.
- [33] S. Shao, Z. Zhao, B. Li et al., "CrowdHuman: a benchmark for detecting human in a crowd," 2018, <https://arxiv.org/abs/1805.00123>.
- [34] T. Y. Lin, M. Maire, S. Belongie et al., "Microsoft COCO: common objects in context," in *Proceedings of the European*

*Conference on Computer Vision*, Springer, Zurich, Switzerland, pp. 740–755, September 2014.

- [35] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the Computer Vision and Pattern Recognition*, IEEE, LasVegas, NV, USA, pp. 770–778, June 2016.
- [36] J. Hosang, R. Benenson, P. Dollár, and B. Schiele, “What makes for effective detection proposals?” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 4, pp. 814–830, 2016.

## Research Article

# Hybrid Depth-Separable Residual Networks for Hyperspectral Image Classification

Cuijie Zhao,<sup>1,2</sup> Hongdong Zhao ,<sup>1</sup> Guozhen Wang,<sup>3</sup> and Hong Chen<sup>1</sup>

<sup>1</sup>College of Electronic Information Engineering, Hebei University of Technology, Tianjin 300401, China

<sup>2</sup>Tianjin University of Finance and Economics, Pearl River College, Tianjin 301811, China

<sup>3</sup>Department of Computer Science and Technology, Tianjin University Renai College, Tianjin 300000, China

Correspondence should be addressed to Hongdong Zhao; zhaohd@hebut.edu.cn

Received 13 July 2020; Revised 10 August 2020; Accepted 12 August 2020; Published 26 August 2020

Academic Editor: Zhihan Lv

Copyright © 2020 Cuijie Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

At present, the classification of the hyperspectral image (HSI) based on the deep convolutional network has made great progress. Due to the high dimensionality of spectral features, limited samples of ground truth, and high nonlinearity of hyperspectral data, effective classification of HSI based on deep convolutional neural networks is still difficult. This paper proposes a novel deep convolutional network structure, namely, a hybrid depth-separable residual network, for HSI classification, called HDSRN. The HDSRN model organically combines 3D CNN, 2D CNN, multiresidual network ROR, and depth-separable convolutions to extract deeper abstract features. On the one hand, due to the addition of multiresidual structures and skip connections, this model can alleviate the problem of over fitting, help the backpropagation of gradients, and extract features more fully. On the other hand, the depth-separable convolutions are used to learn the spatial feature, which reduces the computational cost and alleviates the decline in accuracy. Extensive experiments on the popular HSI benchmark datasets show that the performance of the proposed network is better than that of the existing prevalent methods.

## 1. Introduction

HSI has been widely used in environmental monitoring [1], mineral exploration [2], agricultural remote sensing [3], vegetation ecology [4], ocean remote sensing [5], and other earth observation tasks. In these applications, because HSI exhibits mixed land cover categories, resulting in high intraclass variability and interclass similarity, it is a huge challenge for any classification model. In order to improve the performance of HSI classification, traditional machine learning methods integrate spectral features and spatial features to achieve effective feature extraction such as random forest [6], SVM and its variants [7, 8], sparse self-representation [9], and artificial neural network [10]. However, these methods only extract surface features such as edges and textures of HSI, which will reduce the feature representation ability of hyperspectral images.

Deep learning [11, 12] method has been widely used in image processing, especially image classification [13] and

target recognition [14]. It can actively learn to extract features, with little manual intervention, and automatically find effective features. The deep model can also extract high-level abstract features by adding hierarchical abstractions, which are usually more robust to nonlinear processing. The basic network frameworks for deep learning include unsupervised neural networks [15], convolutional neural networks [16], cyclic neural networks [17], and recursive neural networks [18]. Among them, convolutional neural networks (CNNs) are classic models. This model is based on big data and deep network structures. It extracts rich deep features from the original hyperspectral data, ensuring the integrity of spatial and spectral information and avoiding the initiative and randomness of human feature extraction, and it can achieve better classification results than other deep learning models. In addition, the deep convolutional networks AlexNet [19], VGG [20], GoogleNet [21], ResNet [14], etc., perform well, which fully demonstrates the fact that convolutional neural networks are a good strategy for image classification.



In recent years, the classification of HSI based on the deep convolutional network has also made great progress. HSI is 3D data, which include 2D spatial information and 1D spectral information [22]. The classification of HSI based on CNN mainly uses 2D CNN or 3D CNN for hierarchical feature extraction. Chen et al. proposed to use 3D convolution to get spatial-spectral depth features [23]. A 3D CNN was designed by Li for spectral space classification [24]. Subsequently, Yang et al. proposed a 3D recurrent CNN [25]. Song et al. designed a deep feature fusion network to solve the hyperspectral classification problem [26], Fang et al. proposed a deep hash neural network [27], and Gong et al. introduced statistical metric methods for HSI spectral-spatial classification [28]. Zhong et al. designed the SSRN structure and introduced identity mapping residual blocks for spectral-spatial feature learning [22]. Liu et al. used 3D CNN and residual connection to construct a 12-layer deep network ( Res-3D CNN ) [29]. Lee and Kwon adopted the residual connection to make the network deeper and used 11 convolution kernels to learn the hierarchical features [30]. Recently, Roy et al. designed a concise model (HybridSN) [31] that combined 3D CNN and 2D CNN to extract spectral features, and by comparing with the most advanced models in the past, the classification effect is more excellent. Cao and Guo further introduced hybrid dilated convolutions (HDC) and the residual block based on SSRN and proposed a new end-to-end hybrid expansion residual deep convolutional network [32]. Wu et al. designed the 3D ResNeXt structure using feature fusion and label smoothing strategies [33].

It can be clearly seen from the above literature that only using 2D CNN for HSI classification cannot extract a good distinguishing feature map from the spectral dimension, and spectral information will be lost. Using only 3D CNN to extract features can enhance the accuracy, but the complexity will grow, and the performance is worse, when dealing with classes with similar textures in many spectral bands. The main reason for the above situation is that HSI is a 3D data image, with spatial dimensions and spectral dimensions. Using only 2D CNN cannot extract feature images with good discriminating ability from the spectral dimension. Similarly, deep 3D CNN complicates the model and greatly increases the amount of calculation. Moreover, many classes with similar textures, using 3D CNN alone, seem to be even worse.

In order to solve the above shortcomings, our model combines 3D CNN and 2D CNN that can fully extract spectral and spatial feature maps and overcome the shortcomings of single 2D CNN and 3D CNN. In addition, deep networks are difficult to train and are prone to problems such as overfitting, gradient disappearance, and gradient explosion. Therefore, it is natural to introduce the residual network in the model [34, 35] because the residual network can be improved by adding layers. With the depth of the network increasing, the operation costs in the model will also increase. Replacing the traditional 2D convolution with deep separable convolutions [36] can solve the problem of parameter and operation costs and further avoid overfitting. In this study, a new network is proposed by constructing a hybrid deep separable residual network. Firstly, the

framework designs a 3D residual module ROR to extract spatial-spectral mixing features. Subsequently, the feature information is converted from 3D data to a 2D feature map. Finally, the 2D depth-separable convolutions extract spatial features. Depth-separable convolution can enhance the feature learning ability of HSI and reduce the computational complexity. The multiresidual network ROR [37] in the network can enhance the learning ability. In addition, skip connections can extract spatial-spectral mixing features more effectively. The major contributions of this paper are listed as follows:

- (1) A hybrid depth-separable convolution residual network is proposed to enhance the feature learning ability of HSI. In this network, spatial-spectrum 3D CNN and spatial 2D CNN are combined into the model, which can better study deep-level spatial-spectral features.
- (2) We embed multiresidual network ROR in the 3D convolutional layer and the 2D convolutional layer that can greatly decrease the number of parameters, thereby simplifying the network structure and promoting the extraction of deep features.
- (3) In the 2D processing part, the use of deep separable convolutional layers reduces the number of parameters and avoids overfitting.

The rest of the article is organized as follows: in Section 2, we introduce the proposed framework. In Section 3, the HSI set and network configuration are explained, with experimental results and analyses. The conclusions and future directions are given in Section 4.

## 2. Methodology

*2.1. Proposed Model.* The HSI data can be regarded as a 3D cube, where the width of the cube is defined as  $W$ , the height is  $H$ , the spectral band is  $D$ , and the original input of hyperspectral data can be expressed as  $R^{W \times H \times D}$ . Each pixel contains  $D$  spectral measurements and forms a one-hot label vector  $C = (c_1, c_2, \dots, c_n) \in R^{1 \times 1 \times n}$ , where  $n$  represents the land cover categories. We design a HDSRN structure for HSI classification, including six parts, namely, PCA dimensionality reduction, 3D spectral and spatial mixed feature learning process, 3D to 2D deformed part, 2D learning process, average pooling layer, and FC layer. In Figure 1, the HDSRN classification framework is described in detail using the Indian Pines dataset as an example, where  $W = H = 145$  and  $D = 200$ .

In the traditional 2D convolution operation, convolution is applied to the spatial dimension, and the 2D feature map is obtained. However, the HSI is 3D data, and it is necessary to capture the spectral feature; 2D CNN cannot process spectral information. The 3D CNN kernel can simultaneously extract spectral and spatial features, but it increases the computational complexity. In addition, when classifying a large number of features with similar textures on the spectral band, the performance is poor. In order to overcome the shortcomings of 2D CNN and 3D CNN and make full

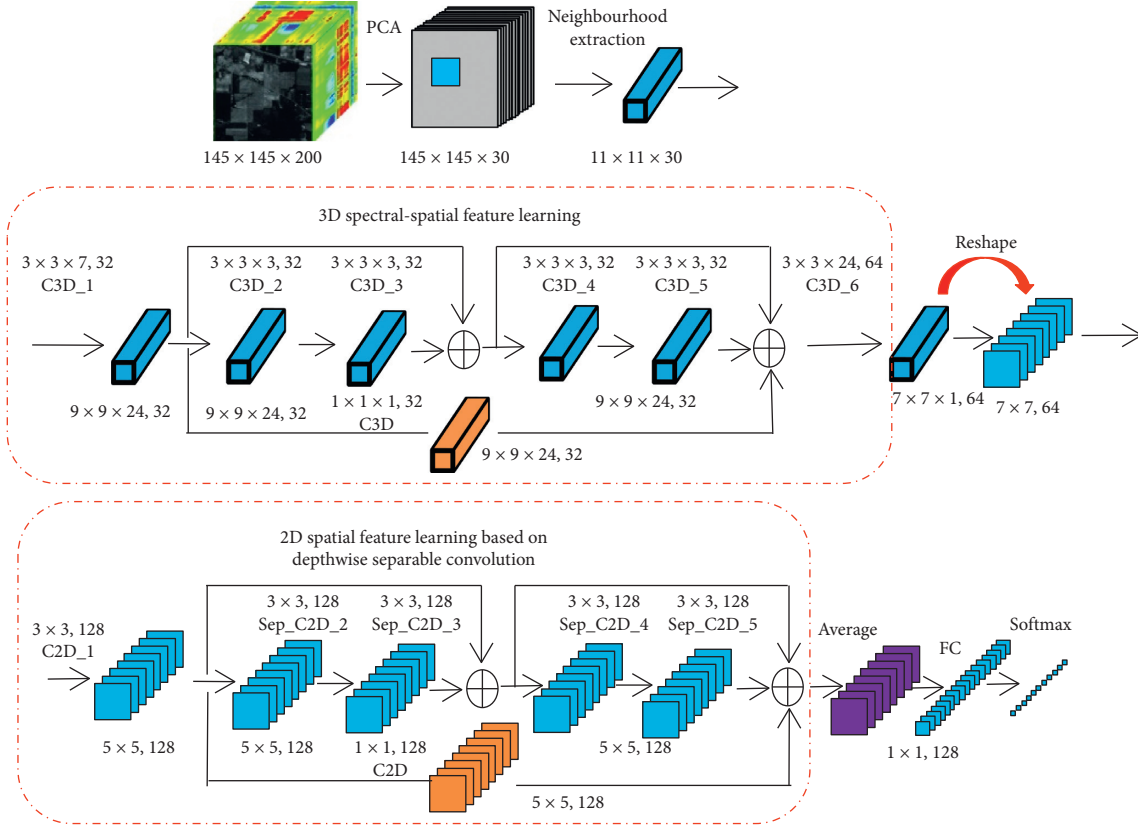


FIGURE 1: The HDSRN classification framework.

use of the automatic feature learning capabilities of 2D and 3D, the model uses a hybrid convolutional neural network framework to combine 3D CNN and 2D CNN into the model. The 3D spatial-spectral feature extraction part consists of two 3D convolutional layers and a set of multi-residual network ROR. The 2D spatial feature is composed of a 2D convolutional layer and a set of multi-residual network ROR. The reshape and average pooling modules are used to adjust the data size to meet the requirements of the next layer. The FC layer and the softmax layer are used to classify HSI.

## 2.2. PCA Dimension Reduction and Data Preprocessing.

In order to remove spectral redundancy, based on the lightweight design, we use PCA to reduce the dimensionality of the original hyperspectral data, retaining relatively few principal components. In Figure 1, PCA decreases the spectral band from  $D$  to  $B$ , and  $B=30$ , while the spatial dimension remains unchanged. In this case, the hyperspectral data are represented as  $R^{W \times H \times B}$ . The HSI image has large size and many bands, and direct processing requires high hardware and memory requirements. Therefore, before image classification and processing, we first divide the HSI image into small pieces. The hyperspectral image is divided into small overlapping 3D patches, expressed as  $R^{S \times S \times B}$ . The height and the width are both  $S$ , and the spectral band is  $B$ . The true value labels are decided by the label of the center pixel. The size of the patches cannot be too small because

being too small will result in a small receiving field and cannot fully extract image features. But, if the patches are too large, the amount of calculation will be large, and the training time and the test time will become longer. Through the comparative experiment in Section 3.2, we find that the classification effect is better when the  $S$  value is 11.

## 2.3. 3D Spectral Space Feature Learning.

HSI data have the characteristics of a spectral-spatial 3D structure. Based on this feature, we construct a 3D convolution network suitable for HSI to extract spatial-spectral features. The 3D convolution operation [38] is achieved by convolving a 3D convolution kernel with 3D data. The input layer is a 3D image, which is composed of a spatial dimension and a spectral dimension. The 3D convolution kernel performs convolution operations on the two dimensions of the input 3D image and obtains a 3D feature map. The 3D convolution formula is as follows:

$$v_{i,j}^{x,y,z} = f \left( \sum_m \sum_{b=0}^{B_i-1} \sum_{h=0}^{H_i-1} \sum_{w=0}^{W_i-1} k_{i,j,m}^{b,h,w} v_{(i-1),m}^{(x+h),(y+w),(z+b)} + b_{i,j} \right), \quad (1)$$

where  $v_{(i-1),m}^{(x+h),(y+w),(z+b)}$  is the value at the position  $(x+h, y+w, z+b)$  of the  $m^{\text{th}}$  feature map output from the  $i-1$  layer,  $k_{i,j,m}^{b,h,w}$  is the value of the  $j^{\text{th}}$  convolution kernel of the  $i^{\text{th}}$  layer at the position  $(b, h, w)$ ,  $B_i$ ,  $H_i$ , and  $W_i$  are the kernel sizes along the spectral and spatial dimensions, respectively,

$(b, h, w)$  is the index of the convolution kernel,  $(x, y, z)$  is the index of the feature map,  $(x, y)$  is the spatial dimension value,  $z$  is the spectral dimension value,  $b_{i,j}$  is the deviation of the  $j^{\text{th}}$  feature map on the  $i^{\text{th}}$  neuron, and  $f$  is the activation function. In the hyperspectral classification task based on 3D convolution, we set the input data to  $W \times H \times B$ ,  $C_1$ , the size of the convolution kernel, is  $k_1 \times k_2 \times k_3$ , and the number of convolution kernels is  $p$ . If there is no padding and the step size is 1 in the convolution operation, the feature map size generated by 3D convolution is  $(W - k_1 + 1) \times (H - k_2 + 1) \times (B - k_3 + 1)$ ,  $p$ . The number of weight parameters of the 3D convolutional layer is  $p \times k_1 \times k_2 \times k_3 \times C_1$ .

Figure 2 is the 3D spatial-spectral feature learning framework. This part is composed of two 3D convolutional layers and a set of multiresidual network ROR. The input data of the network are  $11 \times 11 \times 30$ , the size of the convolution kernel of the first layer is  $3 \times 3 \times 7$ , and the output is 32 feature maps of  $9 \times 9 \times 24$  size. The second layer is a multiresidual network ROR, which uses 3D identity residual blocks to connect to deepen the network, avoiding weak signal loss and excessive fitting, which is conducive to improving efficiency and extracting better deep abstract features without introducing additional parameters. This part uses padding to ensure that the size of the output feature map is the same as the input size. In addition, the outermost skip connection realizes the fusion of feature data by summing the corresponding pixels. Skip connection alleviates the problem of gradient disappearance, contributes to gradient backpropagation, and can more fully extract features. On each convolutional layer connected by the residual block, we use 32 convolution kernels. The size of each kernel is  $3 \times 3 \times 3$ , from which we can get rich spectral and spatial features.

**2.4. 3D to 2D Deformation.** The 2D convolution operation is started, and spatial feature extraction is performed, after the 3D convolution operation. The network outputs 64 feature map data with a size of  $7 \times 7 \times 1$ . In order to learn the output features in the later 2D space, we reshape the 3D features into 64 2D feature maps of size  $7 \times 7$ , as shown in Figure 3. After the reshaping operation, only the 2D spatial features need to be studied, which reduces the parameters and the operation cost compared with 3D convolution.

**2.5. 2D Spatial Feature Learning Based on Depth-Separable Convolution.** Convolutional layer is the core part of CNN to extract deep-level features. The main functional unit of the convolutional layer is a 2D convolution kernel, which acts on the input data of the previous layer to extract features and enhances the model's nonlinear feature extraction capability by adding an activation function, which is beneficial to the extraction of complex deep features. The 2D convolution is shown in the following equation:

$$v_{i,j}^{x,y} v_{(i-1),m}^{(x+h),(y+w)} = f \left( \sum_m \sum_{h=0}^{H_i-1} \sum_{w=0}^{W_i-1} k_{i,j,m}^{h,w} v_{(i-1),m}^{(x+h),(y+w)} + b_{i,j} \right), \quad (2)$$

where  $v_{(i-1),m}^{(x+h),(y+w)}$  is the value at the position  $(x+h, y+w)$  of the  $m^{\text{th}}$  feature map output from the  $i-1$  layer,  $k_{i,j,m}^{h,w}$  is the value of the  $j^{\text{th}}$  convolution kernel of the  $i^{\text{th}}$  layer at the position  $(h, w)$ ,  $B_i$  and  $H_i$  are the kernel sizes along the spatial dimensions,  $(h, w)$  is the index of the convolution kernel,  $(x, y)$  is the index of the feature map,  $b_{i,j}$  is the deviation of the  $j^{\text{th}}$  feature map on the  $i^{\text{th}}$  neuron, and  $f(\cdot)$  is the activation function. In the hyperspectral classification task based on 2D convolution, we set the input data to  $W \times H$ ,  $C_1$ , the size of the convolution kernel is  $k_1 \times k_2$ , and the number of convolution kernels is  $p$ . Through the 2D convolution operation, the current layer can learn local features from the previous layer, and the size of the convolution kernel determines the size of the local space. Continuous convolutional layers can extract deeper and deeper features that are increasingly abstract. Continuous convolutional layer can extract deeper and more abstract features.

Unlike accustomed 2D convolution, depth-separable convolution performs a spatial convolution while keeping the channels independent, and then deep convolution is performed. The depth-separable convolution not only lessens the number of parameters and calculations in the network, but also speeds up the network training speed and reduces the odds of overfitting in HSI classification, as shown in Figure 4.

In general, the input image using convolution operation has the same length and width. Assuming that the input image size is  $H \times H$  and the channel is  $D_1$ , that is, the depth of the image,  $D_2$  convolution kernel size  $h \times h$  convolution operation is used. The specific steps are as follows.

The first step is the deep convolution operation, which uses  $D_1$  convolution kernels of size  $h \times h$  to perform the convolution operation. Each convolution kernel only convolves one channel of the input layer. If no padding is used and the step size is one, the mapping size obtained each time is shown in the following equation:

$$S_M = (H - h + 1)^2, \quad (3)$$

where  $H$  is the length and width of the input image and  $h$  is the convolution kernel size.

These maps are stacked together to create an image, and the size of the image is shown in the following equation:

$$S_h = (H - h + 1)^2 D_1, \quad (4)$$

where  $D_1$  represents the number of channels. Finally, an output image of size  $S_h$  is obtained, and the depth of the image remains the same as the original.

The second step is to expand the depth convolution operation. We use  $D_1$  convolution kernels of size  $1 \times 1$  to perform the operation. Each convolution kernel is convolved with the input image of size  $S_h$  to obtain a size of  $S_M$  mapping. After  $D_2$  times  $1 \times 1$  convolution, the output image can be obtained. The size of the output image is shown in the following equation:

$$I_o = (H - h + 1)^2 D_2. \quad (5)$$

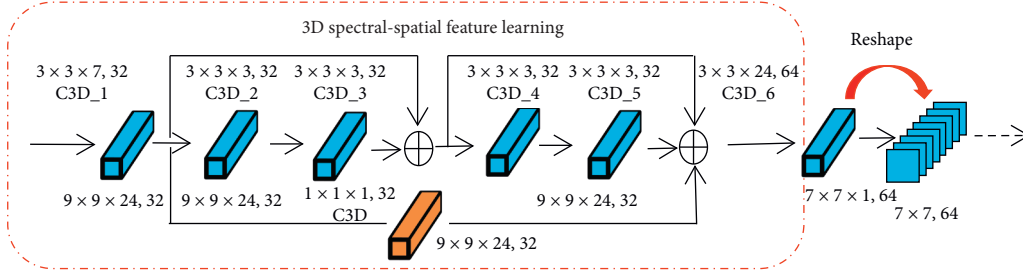


FIGURE 2: 3D spectral-spatial feature learning.

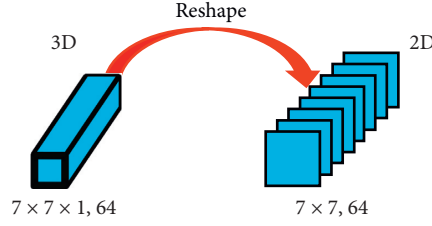


FIGURE 3: Framework of 3D to 2D deformable.

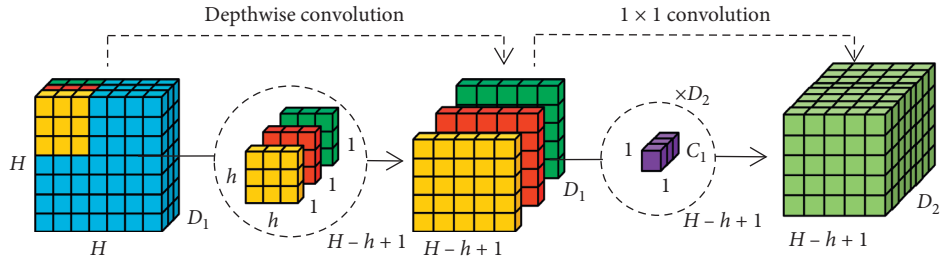


FIGURE 4: Depth-separable convolution.

The depth-separable convolution converts an input layer of  $H \times H \times D_1$  into an output layer of  $I_o$ .

We compare the number of parameters used in the two methods. The number of traditional 2D convolution operation parameters is shown in the following equation:

$$P_t = D_1 h^2 D_2. \quad (6)$$

The number of depth-separable convolution parameters is shown in the following equation:

$$P_d = D_1 h^2 + D_1 D_2, \quad (7)$$

where  $D_1 h^2$  represents the number of operation parameters in the first step and  $D_1 D_2$  is the number of parameters in the second step.

The ratio of the depth-separable convolution and the traditional 2D convolution parameter is shown in equation (8). It can be found that the use of deep separable convolution can greatly reduce the number of parameters and improve operational efficiency:

$$R_p = \frac{1}{D_2} + \frac{1}{h^2}. \quad (8)$$

Figure 5 is a spatial feature learning framework based on depth-separable convolution. This part consists of a

convolutional layer and a set of multiresidual network ROR. In Sep\_C2D\_2, Sep\_C2D\_3, Sep\_C2D\_4, and Sep\_C2D\_5, the depth-separable convolution replaces the traditional 2D convolution operation, which greatly reduces the number of parameters and the operation cost. The input data of the network are 64 feature maps with a size of  $7 \times 7$ . Firstly, this part of the network uses 128 convolution kernels of size  $3 \times 3$  to realize convolution operations on the input data. Secondly, it uses identity residual blocks to connect and deepen the network to extract better deep abstract features. We use padding to keep the size of the output feature map unchanged from the input size. Similarly, the outermost skip connection realizes the fusion of feature data. On each convolutional layer connected by the residual block, we use 128 convolution kernels, and each kernel size is  $3 \times 3$ . Finally, through the average pooling layer, the 128 output feature maps are converted into 128 special maps with a size of  $1 \times 1$ .

**2.6. Residual Network.** In deep learning, the shallow network cannot significantly enhance the classification effect of the network, and the deep network can better learn abstract features [11, 14]. But, the deeper the network, the more obvious the phenomenon of gradient disappearing, and the

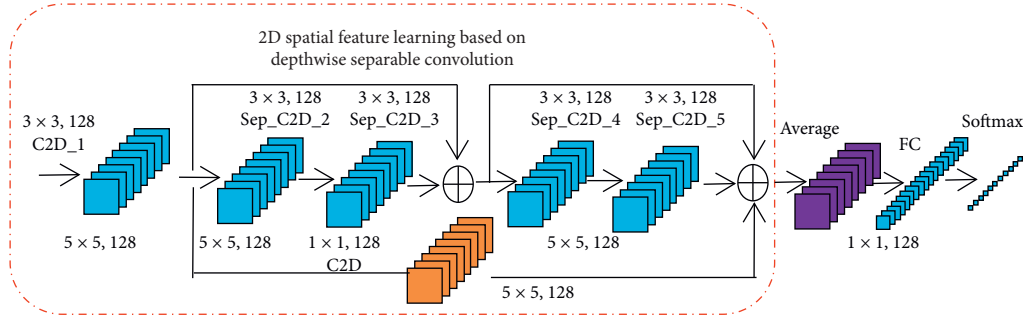


FIGURE 5: 2D spatial feature learning based on depth-separable convolution.

training effect of the network will not be very good. The proposed classification model uses residual connection to deepen the network to solve the gradient dispersion. In addition, the use of residual connections does not introduce additional parameters. According to whether the input and output sizes are the same, ResNet [14] divides the residual connection into identical residual connection and non-identity residual connection, as shown in Figure 6. In the identity residual connection, the input data are  $X$ , and the identity connection is used to inject  $X$  directly into the downstream of the network. The dimension of the output data has not changed, as shown in Figure 6(a). In the nonidentity residual connection, the size of the input data and the output data does not match. A convolutional layer is added to the shortcut path to adjust the input data  $x$  to a suitable size to match the data size of the main path, as shown in Figure 6(b). He et al. [14] show that the identity residual connection can effectively solve network degradation.

Due to the high spectral resolution and high spatial correlation of HSI, we design two consecutive identical residual blocks in the proposed HDSRN model, as shown in Figure 7. In this structure, the gradient in the upper layer can quickly propagate back to the lower layer, thereby facilitating and standardizing the model training process. In addition, we added a skip connection to fuse feature data at the outermost layer of two consecutive identical residual blocks. Skip connection can alleviate the problem of gradient disappearance, help the backpropagation of the gradient, and more fully extract features.

### 3. Experimental Results and Discussion

In this section, we first select three popular HSI benchmark datasets for our experiment and describe evaluation indices and experimental settings. Then, we discuss the impact of input spatial dimensions and proportions of training samples on classification performance. Finally, we compare the HDSRN model with the existing state-of-the-art methods, such as SVM-RBF [39], 2D CNN [40], 3D CNN [24], SSRN [22], and HybridSN [31].

**3.1. Datasets.** We use Indian Pines (IN), University of Pavia (UP), and Salinas Scene (SA) datasets [39] to verify the model.

IN dataset [41, 42] was collected by AVIRIS sensors in northwestern Indiana. The spatial resolution is about 20m, and the image has a  $145 \times 145$  spatial dimension and 224 spectral bands. Among them, 24 spectral bands covering the water absorption area are discarded, so it becomes 3D data of size  $145 \times 145 \times 200$ . The data have 21,025 pixels in total, and the characteristic pixels are 10,249 pixels. The planting area is all agricultural crops with a total of 16 categories.

UP dataset [41, 42] was collected by ROSIS sensors in Pavia University. The spatial resolution is 1.3 m, the image has a  $610 \times 340$  spatial dimension and 103 spectral bands, and its wavelength range is between 430 and 860 nm. The UP dataset contains a total of 2,207,400 pixels, and the feature pixels are only 42,776 and contains 9 categories.

SA dataset [41, 42] was collected by the AVIRIS sensors in Salinas Valley, California, USA. The spatial resolution is 3.7 m, and the image has a  $512 \times 217$  spatial dimension and 224 spectral bands. 20 spectral bands covering the water absorption area are discarded, so it becomes 3D data of size  $512 \times 512 \times 204$ . The SA dataset included a total of 111104 pixels, 56975 pixels are background pixels, and the feature pixels are 54129 pixels with a total of 16 categories.

**3.2. Evaluation Indices.** In this paper, we use the confusion matrix to represent the classification accuracy of HSI. Confusion matrix is an indicator to judge the results of the classification model and is used to judge the quality of the classifier. The judgment indexes customarily used are overall accuracy (OA), average accuracy (AA), and kappa coefficient (kappa).

OA refers to the ratio of the number of correctly classified samples to the overall number of samples, as shown in the following equation:

$$OA = \frac{\sum_{i=1}^n M_{ii}}{N}, \quad (9)$$

where  $N$  is the overall number of samples,  $M_{ii}$  is the diagonal element of the corresponding confusion matrix, and  $n$  is the number of categories.

AA is the average value of the classification accuracy of each category, and its calculation process is shown in the following equation:

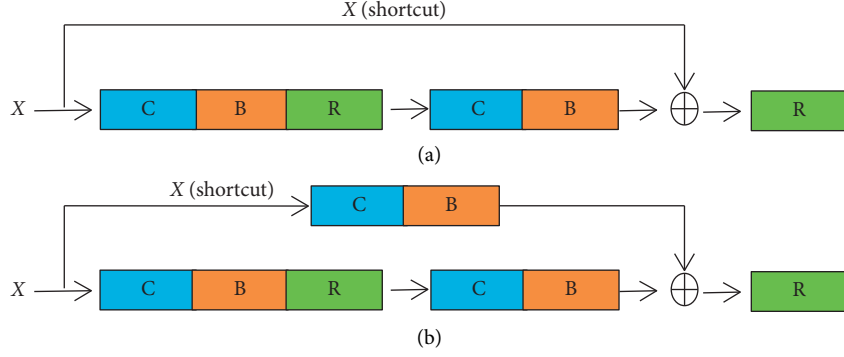


FIGURE 6: The residual connections: (a) identity connection; (b) nonidentity connection.

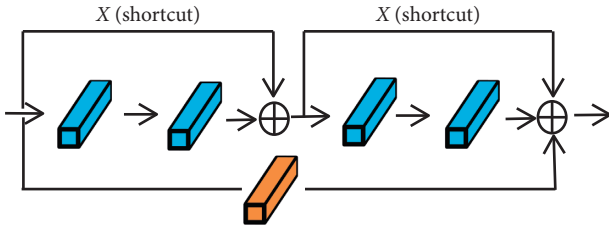


FIGURE 7: Spectral-spatial residual network.

$$AA = \frac{\sum_{i=1}^n \left( M_{ii} / \sum_{j=1}^n M_{ij} \right)}{n} \quad (10)$$

Kappa coefficient is an important indicator to measure classification performance, and its calculation process is shown in the following equation:

$$\text{kappa} = \frac{N \sum_{i=1}^n M_{ii} - \sum_{i=1}^n (M_{i+} + M_{+i})}{N^2 - \sum_{i=1}^n (M_{i+} + M_{+i})} \quad (11)$$

where  $M_{i+}$  is the sum of the  $i^{\text{th}}$  row in the confusion matrix and  $M_{+i}$  represents the sum of the  $i^{\text{th}}$  column in the confusion matrix.

**3.3. Experimental Settings.** We studied the impact of the sample ratio and the input space size on classification performance. For specific experimental details, see Section 3.3. In the end, the most suitable space size we chose is  $11 \times 11$ . For the fairness of comparison, we collected the same spatial dimensions  $11 \times 11$  in the input 3D patches of different datasets. Taking the Indian Pines dataset as an example, the detailed network parameter settings of the model proposed in this paper are shown in Table 1. We applied the Adam algorithm and chose the optimal learning rate of 0.001. The experimental hardware platform is i5-7500 CPU and GTX960 GPU. For these three datasets, in order to get more accurate statistical results, each experiment was repeated 5 times, and the average value of the classification indicators was used as the final result.

**3.4. Experimental Parameter Discussion.** The ratio of training samples and the input space size are two important factors that affect the performance of HSI classification. We use a series of comparative experiments to determine the training sample ratio and the size of the input space. When the training sample ratios of the datasets IN, UP, and SA are 20%, 10%, and 10%, respectively, the classification performance is optimal. Through the experiment in Section 3.4.2, the classification performance is the best when the input space size is  $11 \times 11$ .

**3.4.1. The Impact of the Training Dataset Proportion.** In order to select the appropriate training sample ratio for the dataset, we conducted a comparative experiment with different training sample ratios. The proportions of the training samples we selected are 2%, 5%, 10%, 15%, and 20%. The same spatial dimension is extracted for different datasets. For example, the spatial dimension of IN is  $11 \times 11 \times 30$ , the spatial dimension of UP is  $11 \times 11 \times 15$ , and the spatial dimension of SA is  $11 \times 11 \times 15$ . Each experiment was repeated 5 times, and the average value of the classification indicators was used as observation objects. The changes in OA, AA, and kappa values at different ratios are shown in Table 2. It can be found that as the ratio of input training samples increases, the accuracy gradually improves. When the proportion of training samples in the three datasets reaches 20%, the classification accuracy exceeds 99%, especially the accuracy of the SA dataset is close to 100%. Figure 8 is a classification accuracy chart of three datasets with different training sample proportions. We can clearly see that when the training set is between 2% and 15%, the accuracy is significantly improved. When our training set reaches 15%, the accuracy began to grow slowly. The training sample ratio reaches 20%, and the classification accuracy improves to 99.6%. For the IN dataset, a 20% training sample ratio is enough to train the network. It can be clearly seen from Figures 8(b) and 8(c) that when the training set is between 2% and 10%, the classification accuracy increases significantly. As the training set reaches 10%, the increasing rate of accuracy starts to slow down. The UP classification accuracy reaches 99.73%, and the SA reaches 99.98%. For UP and SA datasets, 10% of the training samples are selected to train the network.

TABLE 1: The architecture of the proposed network.

| Layer           | Output                   | Convolution kernel size | Convolution kernel number |
|-----------------|--------------------------|-------------------------|---------------------------|
| Input           | 11, 11, 30, 1            |                         |                           |
| C3D_1           | 9, 9, 24, 32             | (3, 3, 7)               | 32                        |
| C3D_2           | 9, 9, 24, 32             | (3, 3, 3)               | 32                        |
| C3D_3           | 9, 9, 24, 32             | (3, 3, 3)               | 32                        |
| C3D_4           | 9, 9, 24, 32             | (3, 3, 3)               | 32                        |
| C3D_5           | 9, 9, 24, 32             | (3, 3, 7)               | 32                        |
| C3D             | 9, 9, 24, 32             | (1, 1, 1)               | 32                        |
| C3D_6           | 7, 7, 1, 64              | (3, 3, 24)              | 64                        |
| Reshape         | 7, 7, 64                 |                         |                           |
| C2D_1           | 5, 5, 128                | (3, 3)                  | 128                       |
| Sep_C2D_2       | 5, 5, 128                | (3, 3)                  | 128                       |
| Sep_C2D_3       | 5, 5, 128                | (3, 3)                  | 128                       |
| Sep_C2D_4       | 5, 5, 128                | (3, 3)                  | 128                       |
| Sep_C2D_5       | 5, 5, 128                | (3, 3)                  | 128                       |
| C2D             | 5, 5, 128                | (1, 1)                  | 128                       |
| Average pooling | 1, 1, 128                |                         |                           |
| FC              | 16                       |                         |                           |
| Output          | Categories of land cover |                         |                           |

TABLE 2: Classification accuracy of different training sample ratios.

| Ratios (%) | IN datasets |        |           | UP datasets |        |           | SA datasets |        |           |
|------------|-------------|--------|-----------|-------------|--------|-----------|-------------|--------|-----------|
|            | OA (%)      | AA (%) | Kappa (%) | OA (%)      | AA (%) | Kappa (%) | OA (%)      | AA (%) | Kappa (%) |
| 2          | 86.69       | 82.73  | 85.87     | 98.39       | 96.63  | 97.67     | 99.19       | 98.83  | 99.11     |
| 5          | 94.99       | 95.02  | 93.73     | 99.40       | 98.71  | 98.88     | 99.81       | 99.41  | 99.63     |
| 10         | 98.49       | 98.05  | 98.16     | 99.86       | 99.83  | 99.81     | 99.98       | 99.98  | 99.98     |
| 15         | 99.03       | 99.00  | 98.97     | 99.87       | 99.85  | 99.86     | 99.99       | 99.98  | 99.98     |
| 20         | 99.72       | 99.60  | 99.70     | 99.89       | 99.86  | 99.87     | 100         | 100    | 100       |

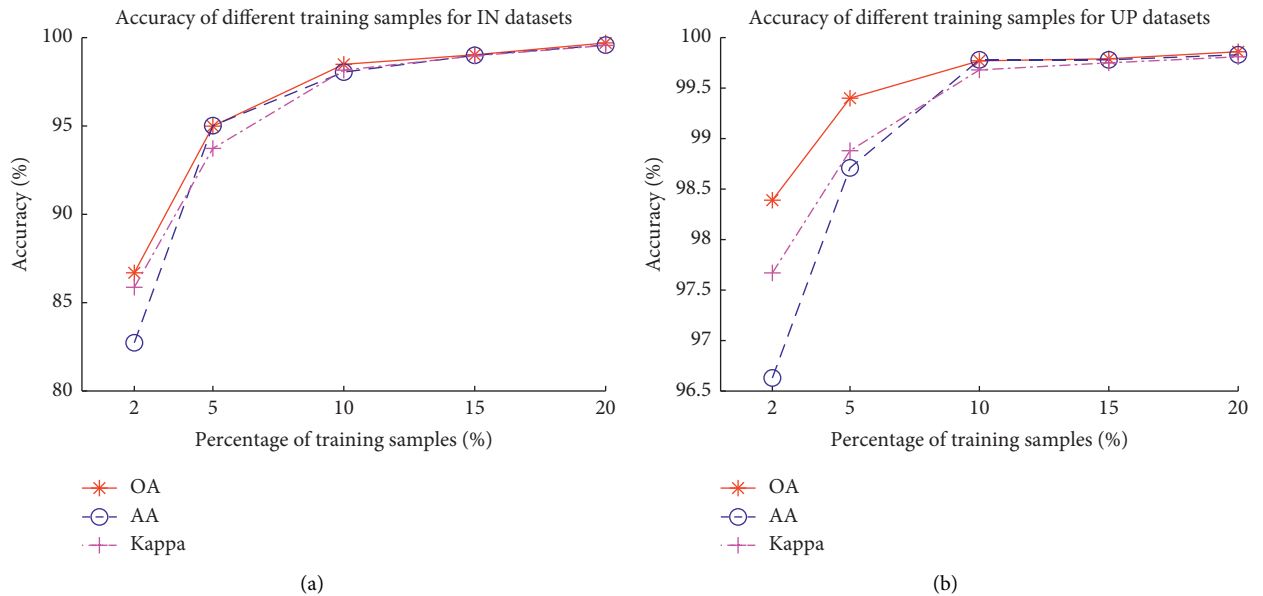
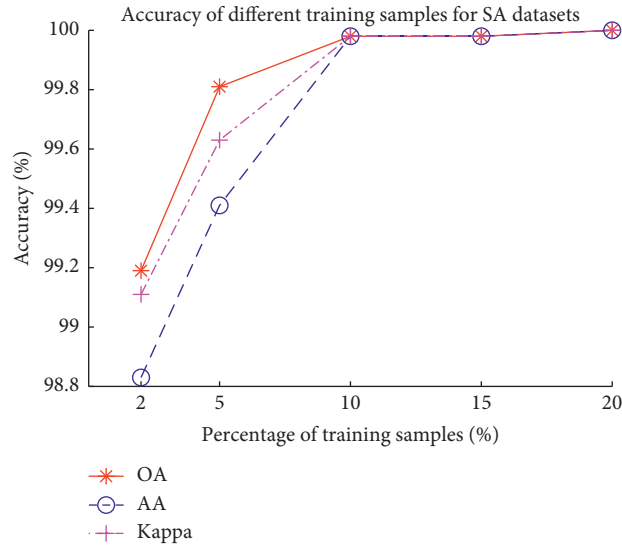


FIGURE 8: Continued.

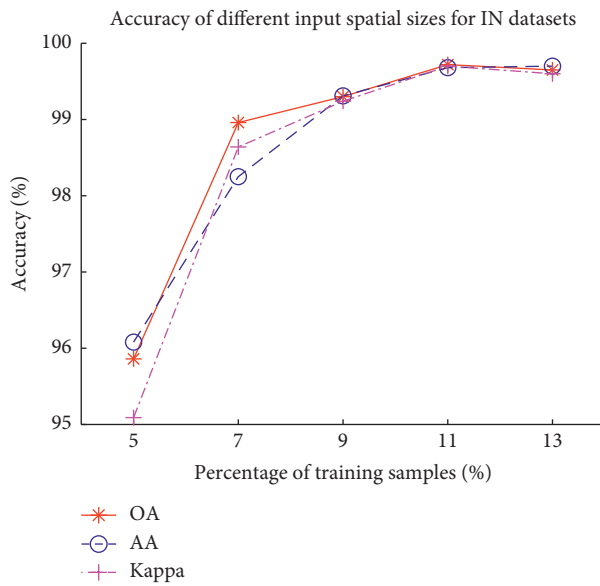


(c)

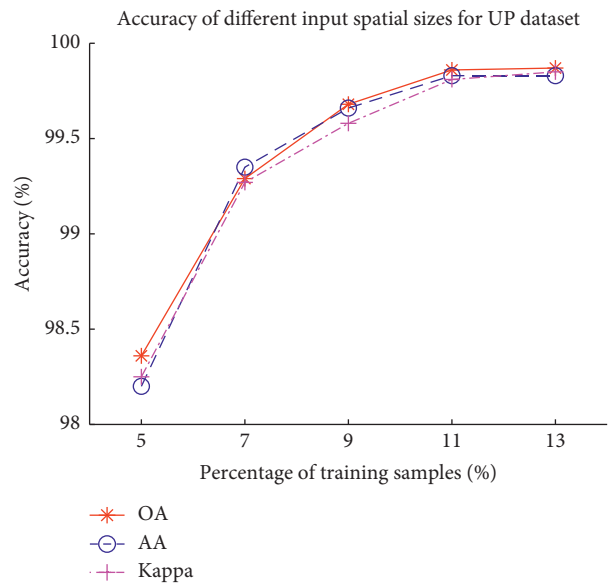
FIGURE 8: Accuracy of different training samples for three datasets: (a) IN; (b) UP; (c) SA.

TABLE 3: Accuracy of different spatial sizes.

| Spatial size | IN datasets |        |           | UP datasets |        |           | SA datasets |        |           |
|--------------|-------------|--------|-----------|-------------|--------|-----------|-------------|--------|-----------|
|              | OA (%)      | AA (%) | Kappa (%) | OA (%)      | AA (%) | Kappa (%) | OA (%)      | AA (%) | Kappa (%) |
| 5 × 5        | 95.86       | 96.08  | 95.09     | 98.36       | 98.20  | 98.25     | 98.89       | 98.88  | 98.85     |
| 7 × 7        | 98.96       | 98.25  | 98.64     | 99.29       | 99.35  | 99.27     | 99.58       | 99.45  | 99.57     |
| 9 × 9        | 99.30       | 99.31  | 99.24     | 99.68       | 99.66  | 99.58     | 99.87       | 99.86  | 99.88     |
| 11 × 11      | 99.72       | 99.60  | 99.70     | 99.86       | 99.83  | 99.81     | 99.98       | 99.98  | 99.98     |
| 13 × 13      | 99.65       | 99.61  | 99.60     | 99.87       | 99.83  | 99.85     | 100         | 100    | 100       |



(a)



(b)

FIGURE 9: Continued.



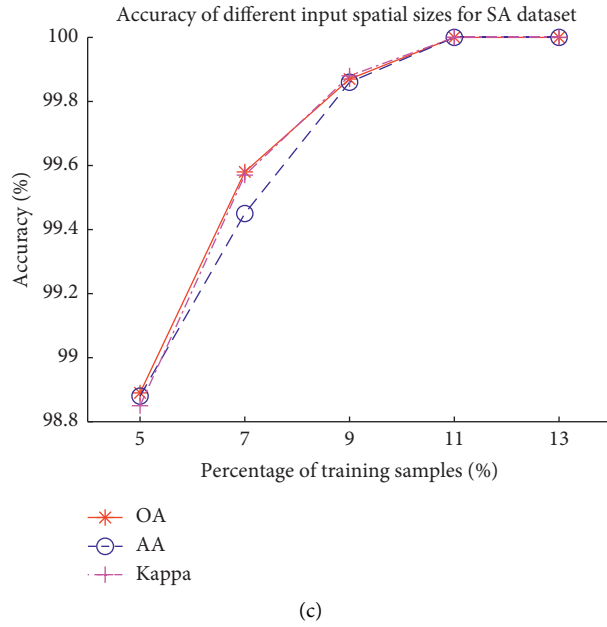


FIGURE 9: Accuracy of different spatial sizes: (a) IN; (b) UP; (c) SA.

TABLE 4: Training time and test time of different spatial sizes.

| Spatial size | IN datasets       |               | UP datasets       |               | SA datasets       |               |
|--------------|-------------------|---------------|-------------------|---------------|-------------------|---------------|
|              | Training time (s) | Test time (s) | Training time (s) | Test time (s) | Training time (s) | Test time (s) |
| 5 × 5        | 94.9              | 1.6           | 99.9              | 2.3           | 100.0             | 3.1           |
| 7 × 7        | 173.5             | 1.9           | 225.7             | 3.5           | 351.1             | 5.2           |
| 9 × 9        | 323.7             | 2.6           | 500.7             | 5.3           | 650.3             | 6.2           |
| 11 × 11      | 577.1             | 3.0           | 787.3             | 6.8           | 926.3             | 8.2           |
| 13 × 13      | 995.1             | 6.2           | 1296.6            | 11.6          | 1537.0            | 14.3          |

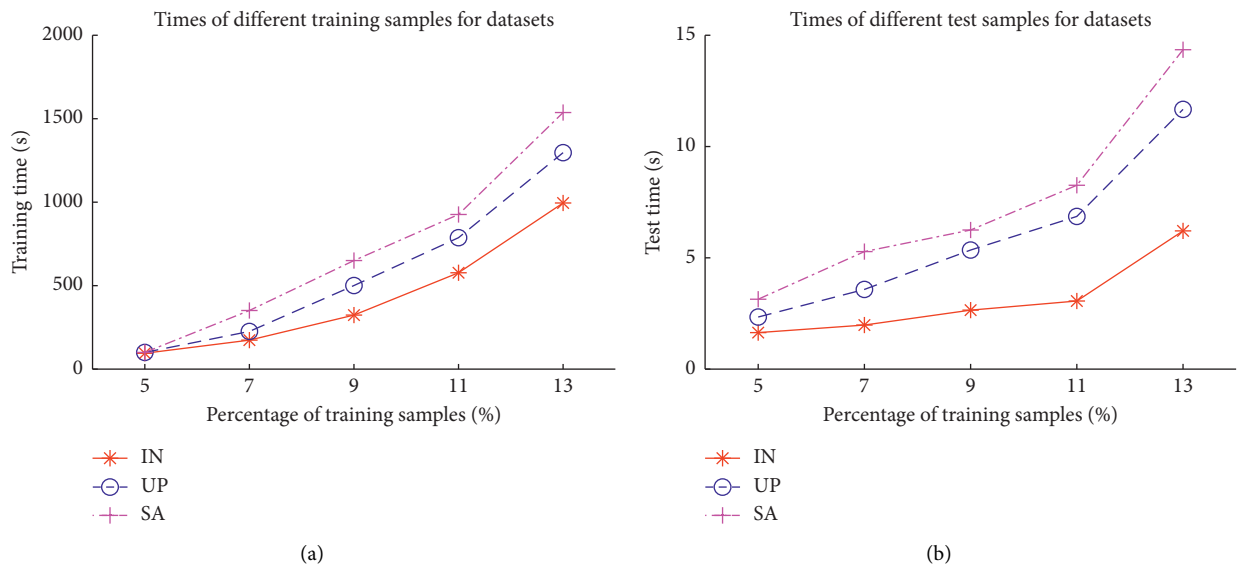


FIGURE 10: (a) Training time and (b) test time of different spatial window sizes for three datasets.

TABLE 5: Accuracy of different residual block combinations.

| Combination | IN datasets |        |           | UP datasets |        |           | SA datasets |        |           |
|-------------|-------------|--------|-----------|-------------|--------|-----------|-------------|--------|-----------|
|             | OA (%)      | AA (%) | Kappa (%) | OA (%)      | AA (%) | Kappa (%) | OA (%)      | AA (%) | Kappa (%) |
| 1 + 1       | 99.48       | 99.58  | 99.40     | 99.73       | 99.70  | 99.72     | 99.85       | 99.78  | 99.82     |
| 1 + 2       | 99.46       | 99.60  | 99.41     | 99.42       | 99.50  | 99.37     | 99.57       | 99.45  | 99.53     |
| 2 + 1       | 99.50       | 99.31  | 99.44     | 99.44       | 99.30  | 99.42     | 99.52       | 99.56  | 99.47     |
| 2 + 2       | 99.72       | 99.60  | 99.70     | 99.86       | 99.83  | 99.81     | 99.98       | 99.98  | 99.98     |
| 2 + 3       | 99.41       | 99.02  | 99.32     | 99.42       | 99.35  | 99.34     | 99.56       | 99.45  | 99.54     |
| 3 + 2       | 98.35       | 97.73  | 98.07     | 99.74       | 99.70  | 99.73     | 99.86       | 99.83  | 99.85     |
| 3 + 3       | 99.72       | 99.51  | 99.65     | 99.86       | 99.84  | 99.83     | 99.98       | 99.95  | 99.96     |

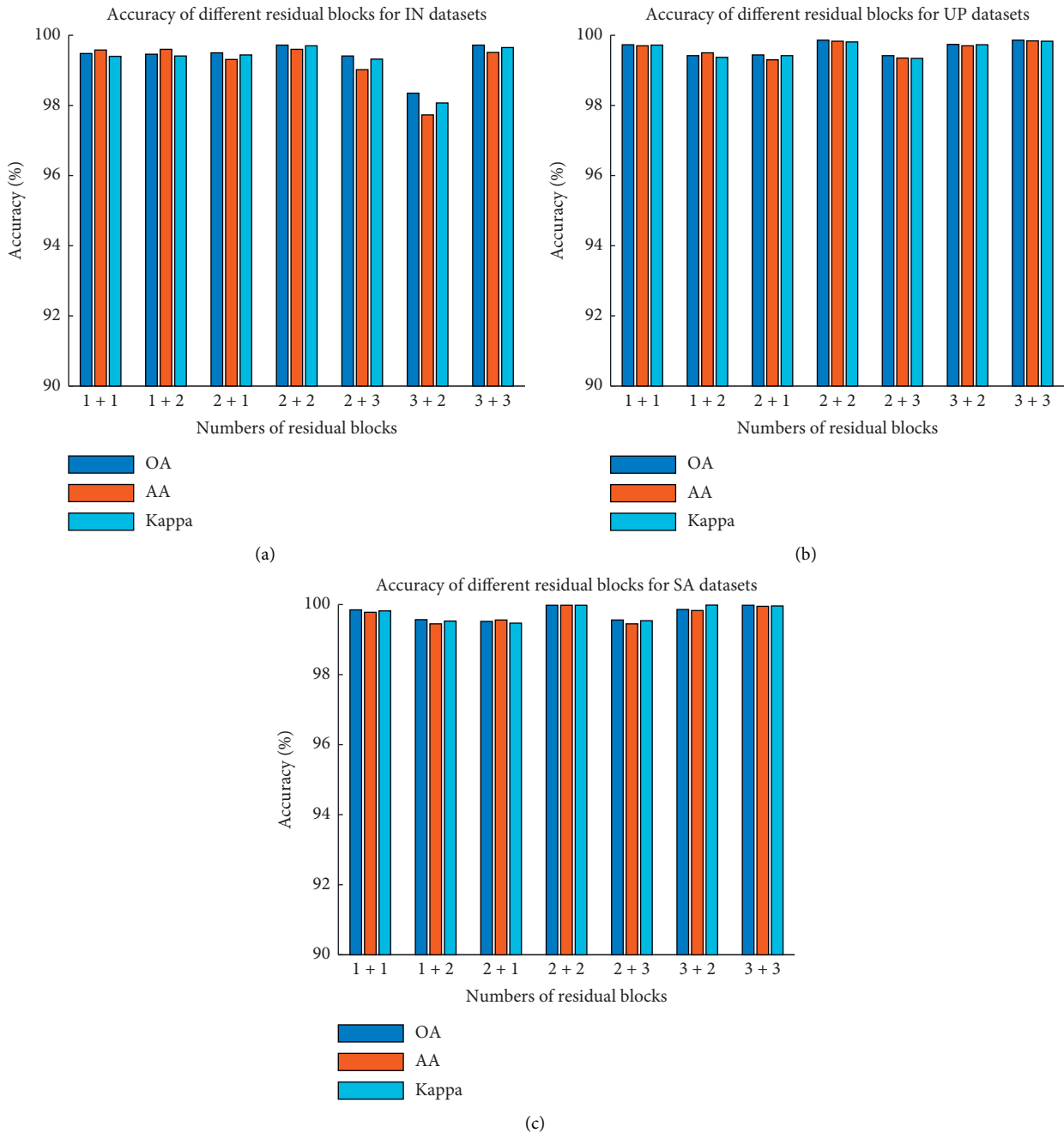


FIGURE 11: Accuracy comparison chart of different combinations: (a) IN; (b) UP; (c) SA.

TABLE 6: Comparative experiment for depth-separable convolution.

| Compare items     | IN datasets |         | UP datasets |         | SA datasets |         |
|-------------------|-------------|---------|-------------|---------|-------------|---------|
|                   | HDSRN       | Model A | HDSRN       | Model A | HDSRN       | Model A |
| Params.           | 714368      | 1233920 | 439904      | 959456  | 439904      | 959456  |
| Training time (s) | 577.1       | 841.8   | 787.3       | 945.7   | 926.3       | 1362.5  |
| Testing time (s)  | 3.0         | 4.5     | 6.8         | 7.5     | 8.2         | 9.5     |
| OA (%)            | 99.72       | 96.24   | 99.86       | 97.36   | 99.98       | 98.55   |

TABLE 7: Classification accuracy of different models.

| Model   | IN datasets |        |           | UP datasets |        |           | SA datasets |        |           |
|---------|-------------|--------|-----------|-------------|--------|-----------|-------------|--------|-----------|
|         | OA (%)      | AA (%) | Kappa (%) | OA (%)      | AA (%) | Kappa (%) | OA (%)      | AA (%) | Kappa (%) |
| HDSRN   | 99.72       | 99.60  | 99.70     | 99.86       | 99.83  | 99.81     | 99.98       | 99.98  | 99.98     |
| Model B | 99.49       | 99.36  | 93.48     | 99.74       | 99.71  | 99.68     | 99.82       | 99.11  | 99.83     |

TABLE 8: The comparison experiments in the IN dataset.

| No. of classes     | Train/test | SVM-RBF | 2D CNN | 3D CNN | SSRN  | HybridSN | HDSRN |
|--------------------|------------|---------|--------|--------|-------|----------|-------|
| 1                  | 9/37       | 61.5    | 85.88  | 94.63  | 97.82 | 97.98    | 99.78 |
| 2                  | 286/1142   | 78.68   | 91.31  | 93.9   | 99.17 | 98.37    | 99.59 |
| 3                  | 166/664    | 73.41   | 91.07  | 94.85  | 99.53 | 99.48    | 99.82 |
| 4                  | 47/190     | 71.58   | 80.38  | 93.48  | 97.79 | 97.38    | 99.82 |
| 5                  | 97/386     | 80.38   | 91.89  | 93.56  | 99.24 | 99.23    | 99.38 |
| 6                  | 146/584    | 92.27   | 99.01  | 94.2   | 99.51 | 99.14    | 99.89 |
| 7                  | 6/22       | 79.52   | 82.59  | 89.73  | 98.7  | 99       | 99.32 |
| 8                  | 96/382     | 87.38   | 100    | 96.01  | 99.85 | 100      | 100   |
| 9                  | 4/16       | 85.87   | 66.55  | 95     | 98.5  | 99.01    | 100   |
| 10                 | 194/778    | 77.58   | 86.38  | 94.55  | 98.74 | 98.76    | 99.61 |
| 11                 | 491/1964   | 83.75   | 90.47  | 93.87  | 99.3  | 99.64    | 99.89 |
| 12                 | 119/474    | 83.21   | 82.89  | 91.52  | 98.43 | 99.12    | 98.55 |
| 13                 | 41/164     | 84.64   | 99.06  | 93.89  | 100   | 100      | 99.86 |
| 14                 | 253/1012   | 98.01   | 97.86  | 91.77  | 99.31 | 100      | 100   |
| 15                 | 77/309     | 94.3    | 90.52  | 95.03  | 99.2  | 99.35    | 99.89 |
| 16                 | 19/74      | 61.43   | 98.94  | 93.57  | 97.82 | 97.26    | 98.57 |
| OA (%)             |            | 82.83   | 90.89  | 94.07  | 99.19 | 99.26    | 99.72 |
| AA (%)             |            | 80.84   | 89.68  | 93.72  | 98.93 | 98.98    | 99.62 |
| Kappa $\times 100$ |            | 82.23   | 88.56  | 93.87  | 99.07 | 99.09    | 99.70 |

TABLE 9: The comparison of experiments in the UP dataset.

| No. of classes     | Train/test | SVM-RBF | 2D CNN | 3D CNN | SSRN  | HybridSN | HDSRN |
|--------------------|------------|---------|--------|--------|-------|----------|-------|
| 1                  | 663/5968   | 93.68   | 97.37  | 97.4   | 99.75 | 99.76    | 99.78 |
| 2                  | 1865/16784 | 97.02   | 99.26  | 94.73  | 99.79 | 99.78    | 99.8  |
| 3                  | 210/1889   | 82.41   | 80.73  | 95.05  | 98.29 | 99.01    | 99.82 |
| 4                  | 306/2758   | 96.51   | 95.54  | 98.04  | 99.52 | 99.53    | 99.56 |
| 5                  | 135/1211   | 98.38   | 99.75  | 99.01  | 99.82 | 99.87    | 99.9  |
| 6                  | 503/4526   | 90.01   | 93.14  | 98.62  | 99.77 | 99.77    | 99.81 |
| 7                  | 133/1197   | 85.92   | 91.65  | 97.02  | 99.65 | 99.69    | 100   |
| 8                  | 368/3314   | 88.08   | 92.39  | 98.23  | 99.05 | 99.21    | 99.8  |
| 9                  | 95/852     | 99.85   | 99.09  | 99.29  | 99.78 | 99.80    | 100   |
| OA (%)             |            | 82.67   | 96.89  | 99.07  | 99.62 | 99.72    | 99.86 |
| AA (%)             |            | 80.84   | 95.79  | 98.75  | 99.49 | 99.60    | 99.83 |
| Kappa $\times 100$ |            | 81.21   | 96.56  | 98.87  | 99.50 | 99.64    | 99.81 |

3.4.2. *The Effect of Space Size.* In the deep convolutional neural network, the larger the size of the input image, the larger the number of operation parameters and the higher the computational complexity. In addition, if the size of the

input image is too small, the receiving field will be too small to obtain a good classification result. In our experiment, we tried 5 different spatial input sizes, namely,  $5 \times 5$ ,  $7 \times 7$ ,  $9 \times 9$ ,  $11 \times 11$ , and  $13 \times 13$ , to evaluate the influence of the network

TABLE 10: The comparison of experiments in the SA dataset.

| No. of classes     | Train/test | SVM-RBF | 2D CNN | 3D CNN | SSRN  | HybridSN | HDSRN  |
|--------------------|------------|---------|--------|--------|-------|----------|--------|
| 1                  | 201/1808   | 92.56   | 95.51  | 95.19  | 99.78 | 100.00   | 100.00 |
| 2                  | 373/3353   | 94.78   | 95.87  | 96.12  | 99.78 | 100.00   | 100.00 |
| 3                  | 198/1778   | 94.22   | 95.64  | 95.89  | 99.78 | 100.00   | 100.00 |
| 4                  | 139/1255   | 97.36   | 99.99  | 97.18  | 99.99 | 100.00   | 100.00 |
| 5                  | 268/2410   | 93.38   | 94.51  | 94.70  | 99.78 | 100.00   | 100.00 |
| 6                  | 396/3563   | 95.36   | 96.71  | 95.63  | 99.78 | 100.00   | 100.00 |
| 7                  | 358/3221   | 95.79   | 95.51  | 95.36  | 99.78 | 100.00   | 100.00 |
| 8                  | 1127/10144 | 80.21   | 86.93  | 86.33  | 99.78 | 100.00   | 100.00 |
| 9                  | 620/5583   | 97.67   | 97.42  | 97.19  | 99.78 | 100.00   | 100.00 |
| 10                 | 328/2950   | 88.99   | 90.93  | 90.94  | 92.69 | 100.00   | 100.00 |
| 11                 | 107/961    | 90.20   | 94.24  | 90.20  | 90.68 | 98.24    | 99.79  |
| 12                 | 193/1734   | 96.42   | 99.91  | 96.24  | 99.95 | 99.61    | 100.00 |
| 13                 | 92/824     | 95.53   | 95.53  | 95.31  | 99.85 | 99.59    | 100.00 |
| 14                 | 107/963    | 91.26   | 92.97  | 94.18  | 99.88 | 100.00   | 100.00 |
| 15                 | 727/6541   | 72.67   | 90.93  | 81.39  | 99.12 | 99.32    | 99.84  |
| 16                 | 181/1626   | 89.37   | 93.43  | 94.09  | 99.78 | 100.00   | 100.00 |
| OA (%)             |            | 92.67   | 95.34  | 94.02  | 99.64 | 99.80    | 99.98  |
| AA (%)             |            | 91.61   | 94.75  | 93.49  | 98.76 | 99.80    | 99.98  |
| Kappa $\times 100$ |            | 92.21   | 94.93  | 93.57  | 99.60 | 99.80    | 99.98  |

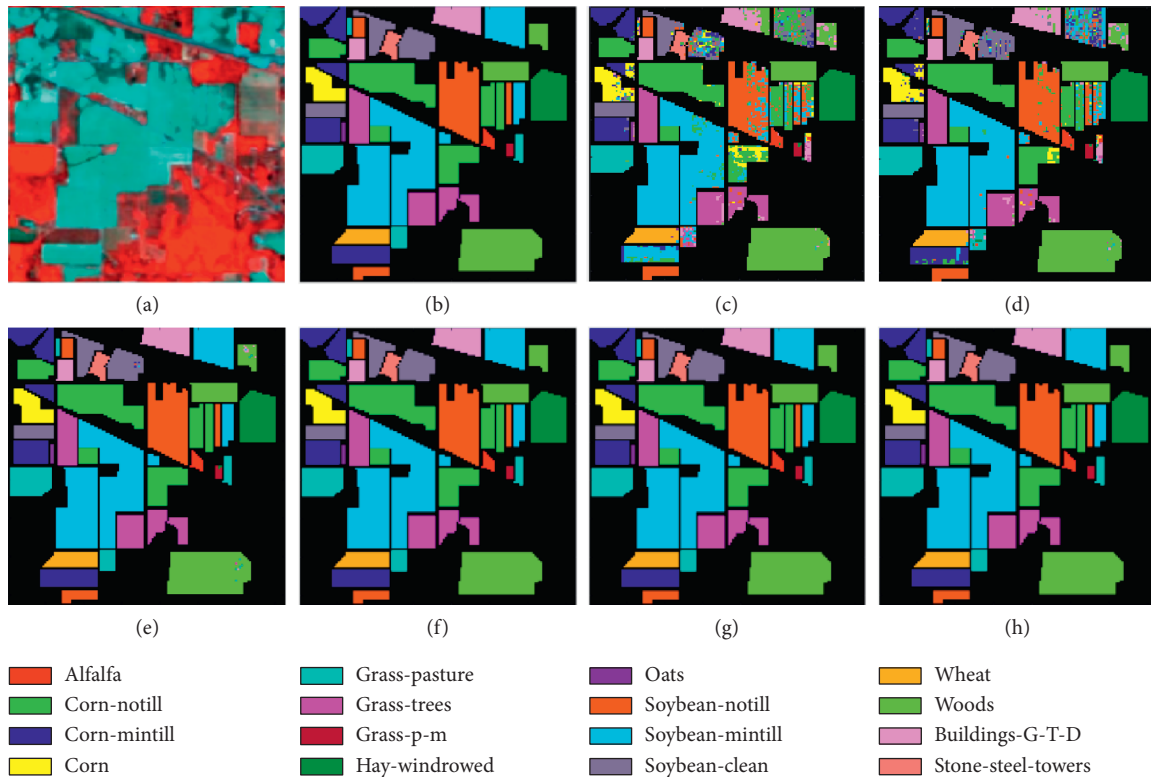


FIGURE 12: Classification results of the models in comparison with the UP dataset. (a) False color image, (b) ground truth, and (c)-(h) predicted classification maps for SVM-RBF, 2D CNN, 3D CNN, SSRN, HybridSN, and proposed HDSRN.

input size. For the IN, UP, and SA datasets, the training dataset ratios are 20%, 10%, and 10%, each experiment was repeated 5 times, and the average value of the classification indicators was used as the observation object. Table 3 and Figure 9 show the changes in OA, AA, and kappa in the IN, UP, and SA datasets under different spatial sizes. Figure 9(a) shows that as the size of the input space increases, the classification accuracy of the IN dataset begins to change

slowly after  $11 \times 11$ . When the space input size is  $13 \times 13$ , there is a slight downward trend. Figure 9(b) and Figure 9(c) show that as the size of the input space gradually increases, the classification accuracy of UP and SA datasets is significantly improved, and when the space input size reaches  $11 \times 11$ , it starts to change slowly.

Table 4 and Figure 10 show the changes in training time and test time under different spatial sizes. We can see from

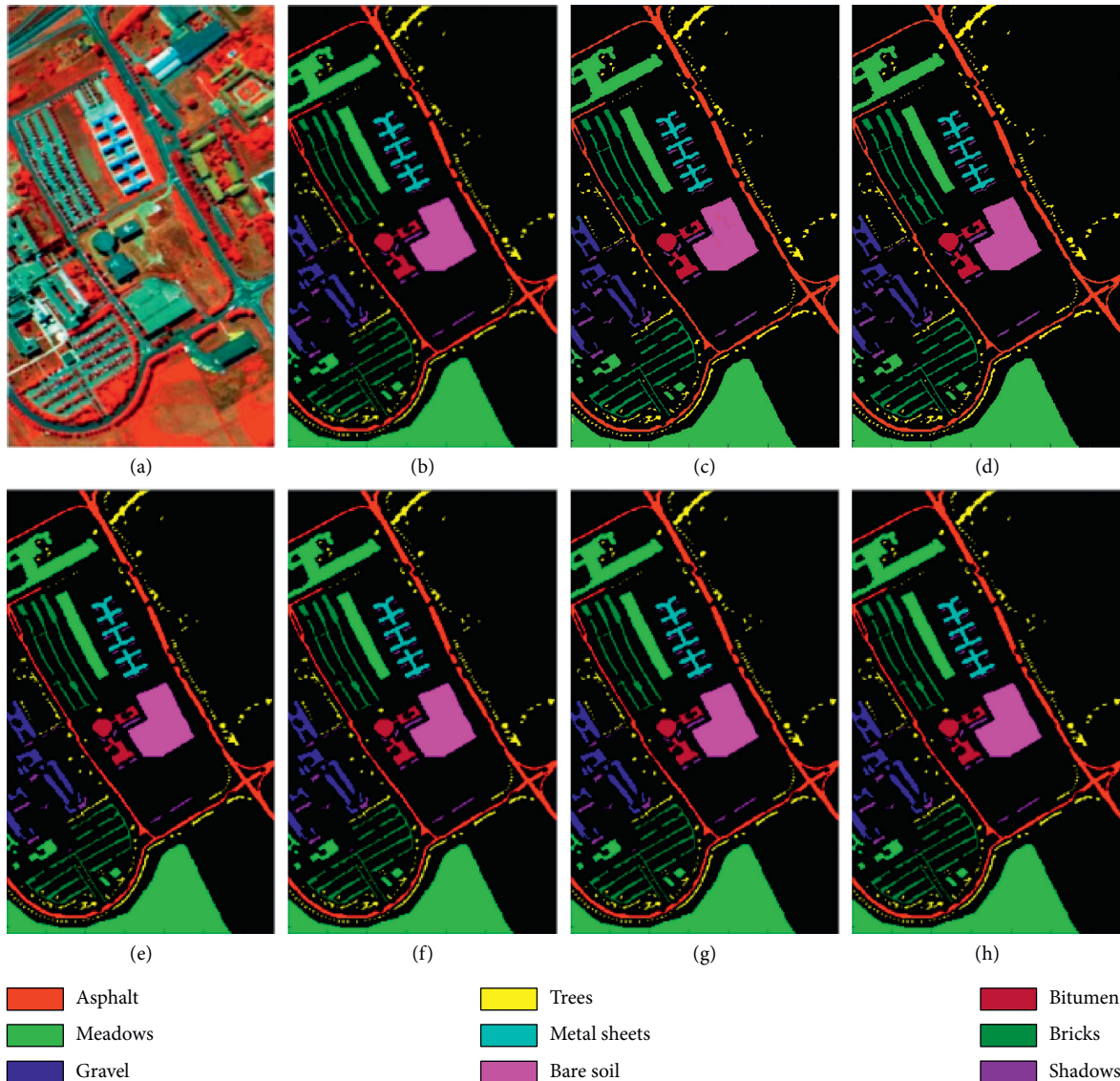


FIGURE 13: Classification maps for the IN dataset. (a) False color image, (b) ground truth, and (c)-(h) predicted classification maps for SVM-RBF, 2D CNN, 3D CNN, SSRN, HybridSN, and proposed HDSRN.

Figure 10 that as the size of the input space increases, the number of calculation parameters also gradually increases, and the training time and test time both increase sharply, resulting in a significant jump in the calculation cost. Through the analysis of the effect of the above space size on the proposed model, we find that using a space size of  $11 \times 11$  is the most suitable for our model. Therefore, the input space size we chose in the comparison experiment is  $11 \times 11$ . At this time, the accuracy of the three datasets is 99.72% (IN), 99.86% (UP), and 99.98% (SA).

**3.4.3. The Effect of Residual Networks.** We conducted a comparative experiment with different combinations of residual blocks. We tested the combination of 3D residual blocks and 2D residual blocks of 1+1, 1+2, 2+1, 2+2, 2+3, 3+2, and 3+3 and gave each group's accuracy, as shown in Table 5. Figure 11 is a comparison bar chart of the

classification results. We can find that the classification effect of the three datasets is the best, when the combination is 2+2. In addition, we can also find from Figure 11 that the accuracy has not continued to improve, but has decreased, as the number of residual blocks increases. The first reason may be limited training samples, and the second reason may be that a deeper network increases the complexity of feature extraction.

**3.4.4. The Effect of Depth-Separable Convolution.** We conducted a comparative experiment to test the impact of depth-separable convolution. The traditional 2D convolution is replaced by the depth-separable convolution to form a comparison model A. The other settings were consistent with HDSRN. The training sample ratios of the IN, UP, and SA datasets are 20%, 10%, and 10%. The same spatial dimension is extracted for different datasets. For example, the

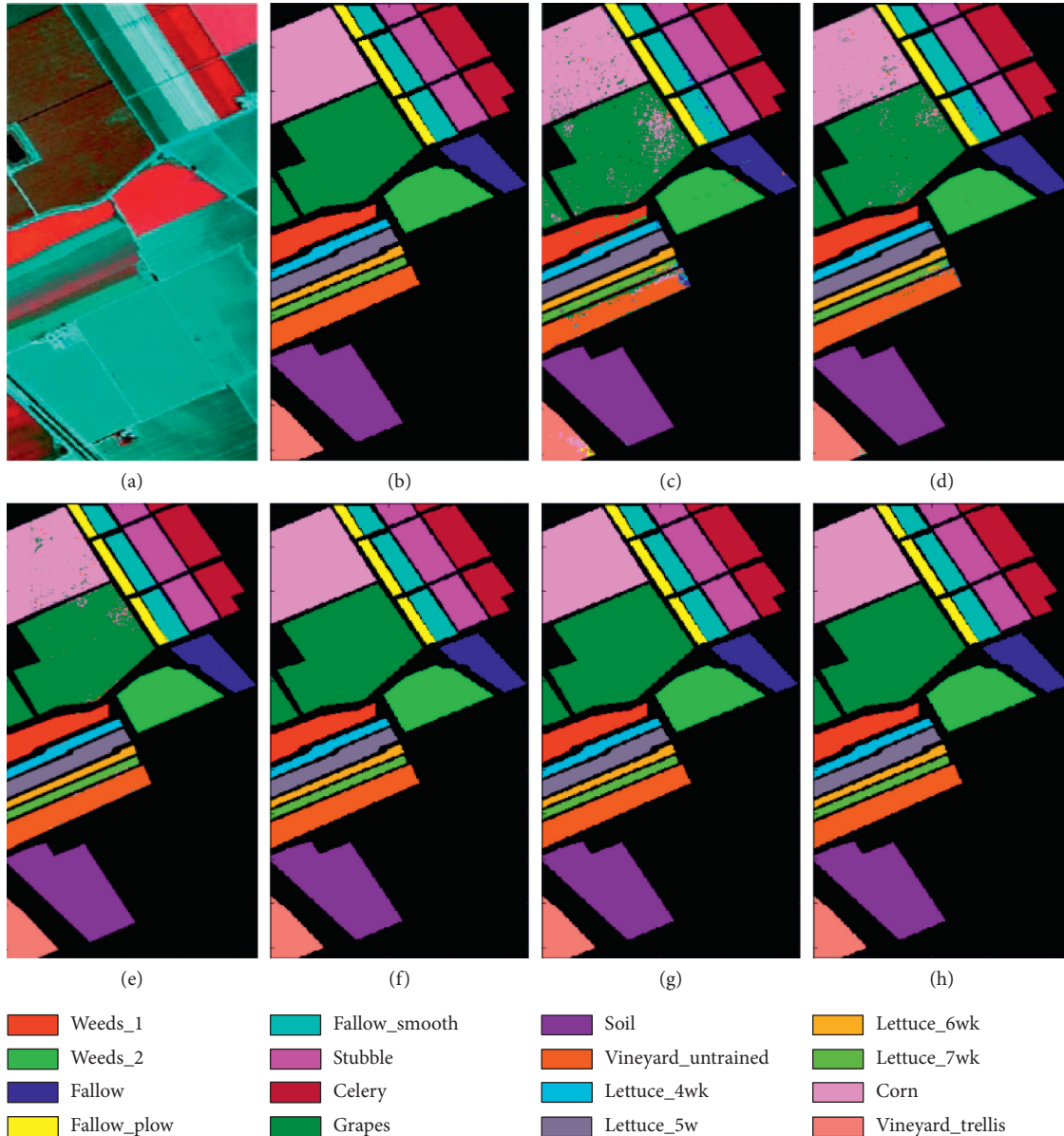


FIGURE 14: Classification maps for the SA dataset. (a) False color image, (b) ground truth, and (c)-(h) classification maps for SVM-RBF, 2D CNN, 3D CNN, SSRN, HybridSN, and proposed HDSRN.

spatial dimension of IN is  $11 \times 11 \times 30$ , the spatial dimension of UP is  $11 \times 11 \times 15$ , and the spatial dimension of SA is  $11 \times 11 \times 15$ . For three datasets, we conducted 5 experiments. Table 6 shows the comparison between traditional 2D convolution and depth-separable convolution in parameters, sample training time, testing time, and overall classification accuracy. We can find that the depth-separable convolution reduces the number of parameters and operation time, avoids overfitting, and improves the performance of HSI classification.

**3.4.5. The Effect of Skip Connections.** The HDSRN framework included a skip layer. In order to test the effect of skip connection, we designed a framework without skip connection as model B and conducted related comparative

experiments on three datasets. We can see from Table 7 that the outermost skip connection can improve the classification accuracy of HSI because skip connection alleviates the problem of gradient disappearance, helps gradients propagate backward, and can extract features more fully.

**3.5. The Comparative Experiment with Popular Methods.** So as to evaluate the HSI classification capability of HDSRN, we compared the model with the popular methods, such as SVM-RBF [37], 2D CNN [38], 3D CNN [24], SSRN [22], and HybridSN [31]. We used some public codes to train and test the data, which can be accessed online at <https://github.com/eecn/Hyperspectral-Classification> and <https://github.com/gokriznastic/HybridSN>. For the fairness of the experiment, through the comparative experiment in Section 3.4,

we set the input space to the same size. The input sizes of IN, UP, and SA datasets are  $11 \times 11 \times 30$ ,  $11 \times 11 \times 15$ , and  $11 \times 11 \times 15$ , respectively. The training data ratios of IN, UP, and SA datasets are 20%, 10%, and 10%, respectively. We conducted 5 repeated trials with the mean classification metrics as final results.

Tables 8-10 show the comparison of experiments using different methods. We can see that the HDSRN method performs well, and the classification effect on the three datasets is better than other methods. One possible reason is that the proposed network model is based on the spatial spectrum 3D CNN and 2D CNN hierarchical framework, and they are complementary. This design method helps to capture more contextual information. Another possible reason is that the multiresidual network ROR is embedded in the proposed model, which can extract deeper abstract features. The last reason is that the depth-separable convolution reduces the computational cost and avoids overfitting.

From the analysis of Tables 8-10, we can see that in the IN dataset, the classification accuracy of HDSRN is about 0.5% higher than SSRN and about 0.4% higher than HybridSN. In the UP and SA datasets, the classification accuracy of HDSRN is also significantly improved. We can also find that the accuracy of network classification that only focuses on spectral or spatial features is usually less than 97%. Therefore, the classification method based on the combination of spectrum and spatial features is significantly better than the traditional method. In the IN dataset, HDSRN obtained significantly better classification results than SSRN and HybridSN, in the classification of categories 2, 5, 9, and 16. In addition, because there are fewer training samples for classes 1, 7, 9, and 16, the classification accuracy of these types of features is unstable and significantly lower than other categories, which has a greater impact on the overall classification accuracy. However, HDSRN can still classify these categories, and the classification accuracy is higher than 97%.

Figures 12-14 are the visualization of the classification by the HDSRN model and the comparison network. These maps include false color images, ground truth, and visualization images of different comparison methods. We can see that SVM-RBF has the worst visual effect among these models. The generated visual image is relatively rough, the classification accuracy is low, and the noise is obvious. This may be due to the fact that traditional methods cannot effectively extract spatial-spectral features, resulting in unsatisfactory classification results. The second is 2D CNN, and the classification effect is relatively poor. This may be because 2D CNN is unable to extract a good identification feature map from the spectral dimension, and a lot of spectral information is lost. The visual images of 3D CNN are relatively smooth. The visual images of SSRN, HybridSN, and HDSRN are smoother. However, SSRN has misclassified noise in categories 7 and 9 of the UP dataset. Compared with other methods, HDSRN has a higher classification accuracy. In addition, the classification map provided by HDSRN is the most accurate, and the edge contours of features are clearer than other methods.

## 4. Conclusions

In this paper, we propose a hybrid separable convolutional residual model for HSI classification. The model integrates spatial-spectral residual blocks, spatially separable convolutional residual blocks, and the outermost skip connections. The spatial-spectral 3D feature and the spatial 2D feature can be continuously extracted. Concretely, the HDSRN model can extract spectral feature information and spatial feature information, and these feature information is complementary. Additionally, we embed a multilevel residual network ROR in the model to learn spectral and spatial features, which improves the network optimization and learning capabilities. Finally, in spatial feature learning, we use separable convolution to extract valuable features, reducing the number of parameters and computing time and alleviating the decline in accuracy. The comparative experimental results on the IN, UP, and SA benchmark datasets verify the superiority of the HDSRN method. The focus of future work will be to use transfer learning methods to further solve the HSI classification problem.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This research was funded by the Foundation of Science and Technology on Electro-Optical Information Security Control Laboratory, grant no. 614210701041705.

## References

- [1] N. Ma, Y. Peng, S. Wang, and P. Leong, "An unsupervised deep hyperspectral anomaly detector," *Sensors*, vol. 18, no. 3, p. 693, 2018.
- [2] G. Notesco, Y. Ogen, and E. Ben-Dor, "Mineral classification of makhtesh ramon in Israel using hyperspectral longwave infrared (LWIR) remote-sensing data," *Remote Sensing*, vol. 7, no. 9, pp. 12282-12296, 2015.
- [3] T.-H. Hsieh and J.-F. Kiang, "Comparison of CNN algorithms on hyperspectral image classification in agricultural lands," *Sensors*, vol. 20, no. 6, p. 1734, 2020.
- [4] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 2, pp. 6-36, 2013.
- [5] D. Prasad and K. Agarwal, "Classification of hyperspectral or trichromatic measurements of ocean color data into spectral classes," *Sensors*, vol. 16, no. 3, p. 413, 2016.
- [6] X. Li, W. Chen, Q. Zhang, and L. Wu, "Building Auto-Encoder Intrusion Detection System based on random forest feature selection," *Computers & Security*, vol. 95, p. 101851, 2020.
- [7] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE*

- Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2094–2107, 2014.
- [8] C. Zhao, H. Zhao, G. Wang, and H. Chen, “Improvement SVM classification performance of hyperspectral image using chaotic sequences in artificial bee colony,” *IEEE Access*, vol. 8, pp. 73947–73956, 2020.
  - [9] P. Hu, X. Liu, Y. Cai, and Z. Cai, “Band selection of hyperspectral images using multiobjective optimization-based sparse self-representation,” *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 3, pp. 452–456, 2019.
  - [10] C. Hernández-Espinosa, M. Fernández-Redondo, and J. Torres-Sospedra, “Some experiments with ensembles of neural networks for classification of hyperspectral images,” *Advances in Neural Networks-ISNN 2004*, pp. 912–917, 2004.
  - [11] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
  - [12] D. Heaven, “Why deep-learning AIs are so easy to fool,” *Nature*, vol. 574, no. 7777, pp. 163–166, 2019.
  - [13] M. H. Hesamian, S. Mashohor, M. I. Saripan, and W. A. Wan Adnan, “Effect of image resolution on intensity based scene illumination classification using neural network,” *The Imaging Science Journal*, vol. 63, no. 8, pp. 433–439, 2015.
  - [14] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, June 2016.
  - [15] Y. Wei, X. Luo, L. Hu, Y. Peng, and J. Feng, “An improved unsupervised representation learning generative adversarial network for remote sensing image scene classification,” *Remote Sensing Letters*, vol. 11, no. 6, pp. 598–607, 2020.
  - [16] J. Gu, Z. Wang, J. Kuen et al., “Recent advances in convolutional neural networks,” *Pattern Recognition*, vol. 77, pp. 354–377, 2018.
  - [17] P. Ghamisi, N. Yokoya, J. Li et al., “Advances in hyperspectral image and signal processing: a comprehensive overview of the state of the art,” *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 37–78, 2017.
  - [18] L. Dong, F. Wei, K. Xu, S. Liu, and M. Zhou, “Adaptive multi-compositionality for recursive neural network models,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 3, pp. 422–431, 2016.
  - [19] S. A. Amirshahi, M. Pedersen, and S. X. Yu, “Image quality assessment by comparing CNN features between images,” *Journal of Imaging Science and Technology*, vol. 60, p. 60410, 2016.
  - [20] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” <http://arxiv.org/abs/1409.1556>.
  - [21] C. Szegedy, W. Liu, Y. Jia et al., “Going deeper with convolutions,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, June 2015.
  - [22] Z. Zhong, J. Li, Z. Luo, and M. Chapman, “Spectral-spatial residual network for hyperspectral image classification: a 3-D deep learning framework,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 847–858, 2018.
  - [23] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, “Deep feature extraction and classification of hyperspectral images based on convolutional neural networks,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 6232–6251, 2016.
  - [24] Y. Li, H. Zhang, and Q. Shen, “Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network,” *Remote Sensing*, vol. 9, no. 1, p. 67, 2017.
  - [25] X. Yang, Y. Ye, X. Li, R. Y. K. Lau, X. Zhang, and X. Huang, “Hyperspectral image classification with deep learning models,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 9, pp. 5408–5423, 2018.
  - [26] W. Song, S. Li, L. Fang, and T. Lu, “Hyperspectral image classification with deep feature fusion network,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 6, pp. 3173–3184, 2018.
  - [27] L. Fang, Z. Liu, and W. Song, “Deep hashing neural networks for hyperspectral image feature extraction,” *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 9, pp. 1412–1416, 2019.
  - [28] Z. Gong, P. Zhong, W. Hu, Z. Xiao, and X. Yin, “A novel statistical metric learning for hyperspectral image classification,” <http://arxiv.org/abs/1905.05087>.
  - [29] B. Liu, X. Yu, P. Zhang, and X. Tan, “Deep 3D convolutional network combined with spatial-spectral features for hyperspectral image classification,” *Acta Geodaetica et Cartographica Sinica*, vol. 48, pp. 53–63, 2019.
  - [30] H. Lee and H. Kwon, “Going deeper with contextual CNN for hyperspectral image classification,” *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4843–4855, 2017.
  - [31] S. K. Roy, G. Krishna, S. R. Dubey, and B. B. Chaudhuri, “HybridSN: exploring 3-D-2-D CNN feature hierarchy for hyperspectral image classification,” *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 2, pp. 277–281, 2020.
  - [32] F. Cao and W. Guo, “Deep hybrid dilated residual networks for hyperspectral image classification,” *Neurocomputing*, vol. 384, pp. 170–181, 2020.
  - [33] P. Wu, Z. Cui, Z. Gan, and F. Liu, “Three-dimensional ResNeXt network using feature fusion and label smoothing for hyperspectral image classification,” *Sensors*, vol. 20, no. 6, p. 1652, 2020.
  - [34] G. Li, H. Tang, Y. Sun et al., “Hand gesture recognition based on convolution neural network,” *Cluster Computing*, vol. 22, no. S2, pp. 2719–2729, 2019.
  - [35] W. Cheng, Y. Sun, G. Li, G. Jiang, and H. Liu, “Jointly network: a network based on CNN and RBM for gesture recognition,” *Neural Computing and Applications*, vol. 31, no. S1, pp. 309–323, 2019.
  - [36] F. Chollet, “Deep learning with depthwise separable convolutions,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, July 2017.
  - [37] K. Zhang, M. Sun, T. X. Han, X. Yuan, L. Guo, and T. Liu, “Residual networks of residual networks: multilevel residual networks,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, pp. 1303–1314, 2017.
  - [38] S. Ji, W. Xu, M. Yang, and K. Yu, “3D convolutional neural networks for human action recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 221–231, 2013.
  - [39] B. Kuo, H. Ho, C. Li, C. Hung, and J. Taur, “A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification,” *Applied Earth Observations and Remote Sensing*, vol. 7, pp. 317–326, 2014.
  - [40] Q. Wang, J. Gao, and Y. Yuan, “A joint convolutional neural networks and context transfer for street scenes labeling,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 5, pp. 1457–1470, 2018.
  - [41] F. Luo, B. Du, L. Zhang, L. Zhang, and D. Tao, “Feature learning using spatial-spectral hypergraph discriminant analysis for hyperspectral image,” *IEEE Transactions on Cybernetics*, vol. 49, pp. 2406–2419, 2018.
  - [42] Computational Intelligence Group of the Basque University (UPV/EHU), *Hyperspectral Remote Sensing Scenes*, Computational Intelligence Group of the Basque University (UPV/EHU), San Sebastian, Spain, 2020, [http://www.ehu.es/ccwintco/index.php/Hyperspectral\\_Remote\\_Sensing\\_Scenes](http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes).



## Research Article

# Modeling and Simulation of English Speech Rationality Optimization Recognition Based on Improved Particle Filter Algorithm

Hui Dong 

*Department of Foreign Languages, Tangshan Normal University, Tangshan 063000, Hebei Province, China*

Correspondence should be addressed to Hui Dong; maymayts@tstc.edu.cn

Received 18 May 2020; Revised 24 July 2020; Accepted 3 August 2020; Published 24 August 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Hui Dong. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As one of the most important communication tools for human beings, English pronunciation not only conveys literal information but also conveys emotion through the change of tone. Based on the standard particle filtering algorithm, an improved auxiliary traceless particle filtering algorithm is proposed. In importance sampling, based on the latest observation information, the unscented Kalman filter method is used to calculate each particle estimate to improve the accuracy of particle nonlinear transformation estimation; during the resampling process, auxiliary factors are introduced to modify the particle weights to enrich the diversity of particles and weaken particle degradation. The improved particle filter algorithm was used for online parameter identification and compared with the standard particle filter algorithm, extended Kalman particle filter algorithm, and traceless particle filter algorithm for parameter identification accuracy and calculation efficiency. The topic model is used to extract the semantic space vector representation of English phonetic text and to sequentially predict the emotional information of different scales at the chapter level, paragraph level, and sentence level. The system has reasonable recognition ability for general speech, and the improved particle filter algorithm evaluation method is further used to optimize the defect of the English speech rationality and high recognition error rate. Related experiments have verified the effectiveness of the method.

## 1. Introduction

English speech rationality recognition is the product of the combination of emotion computing and speech synthesis. With the help of the concept of emotion computing, the relationship between emotion and speech is analyzed from the speech signal carrying the known emotional state, and these emotional features are applied to the speech synthesis process. In order to obtain natural and friendly synthesized speech with rich tone changes, which can simulate human emotions, current speech synthesis technology usually refers to text-to-speech conversion technology, which mainly solves how to convert text information into audible sound information. English speech interaction is a natural and convenient way for humans to communicate with machines in the future. Research on speech recognition is enough to promote this technology to serve humans faster and better.

Whether it is from the perspective of technology accumulation or data collection, the current period is a favorable period for studying speech recognition technology.

Particle filter algorithm [1, 2] is an online nonlinear identification algorithm based on Bayesian estimation and Monte Carlo method. Its essence is to approximate the state probability density function by finding a set of random samples propagating in the state space. A discrete sample is used to simulate a continuous function, and the sample mean is used to replace the integral operation, thereby obtaining the process of minimum variance distribution of the state. The particle filtering algorithm theoretically has higher recognition accuracy than the extended Kalman filter algorithm. The authors of [3, 4] improved the traditional particle filter algorithm and used the latest observation information in the importance sampling process to more accurately approximate the posterior probability density

function. The authors of [5, 6] proposed a resampling algorithm to solve the problem of particle degradation. The authors of [7–9] use particle filtering algorithm to solve the damage identification problem of structural systems. Research shows that, compared with EKF algorithm, particle filtering has higher structural model parameter recognition accuracy under non-Gaussian noise conditions. The authors of [10–12] use a hybrid Gaussian particle filter to predict and analyze the posterior distribution parameters and monitor values of the monitoring information state variables in one step forward. At present, the research and application of particle filters in civil engineering are still very limited. How to further improve the sampling accuracy of particle filter algorithms and weaken particles is still the key issue to improve the accuracy of the algorithm. The existing methods for synthesizing emotional speech mainly include three categories: waveform stitching, speech conversion, and statistical parameter synthesis. The waveform stitching method collects voices of different emotion types by recording a large-scale emotional corpus and extracts corresponding speech fragments from the corresponding emotional corpus during synthesis and obtains voices that retain the original recorded tone through stitching [13–15]. The voice conversion method analyzes the change of acoustic parameters of voices of different emotion types relative to neutral voices and adjusts and converts presynthesized neutral voices to obtain new emotional voices [16, 17]. Statistical parameter synthesis is based on statistical models such as Hidden Markov Models. It performs parametric characterization and acoustic modeling of speech with different emotions. Based on this, it performs acoustic prediction of emotional speech and synthesizes speech with different emotions [18–20]. At present, the three methods have their own advantages and disadvantages. The emotional speech that is properly combined by the waveform stitching method is better than other methods. However, the types of emotions that can be synthesized are limited by the existing emotion types of the emotional corpus. The cost of large-scale database building is high; English speech rationality optimization conversion method relies on the study of emotional acoustic feature analysis [21–23]. Due to the diversity and complexity of emotional performance, only some specific emotional states and directional cues were associated with changes in acoustic parameters: Statistical parameter synthesis methods can automatically build a new synthesis system in a short period of time, basically without human intervention, and require less data compared to waveform stitching methods and synthetic emotion types. It is more flexible than the previous two methods, but the spectrum and prosody model generated by HMM is too smooth, which makes the details of the spectrum and prosody model lost, affecting the naturalness of the English speech sound optimization and recognition. With the tremendous achievements of neural networks in many fields, many scholars have begun to study their applications in the field of speech recognition. A variety of models such as deep neural networks, convolutional neural networks, and recurrent neural networks have been introduced and achieved good results [24–26]. The system based on the state output of the deep neural network has a certain decrease in error rate compared

to the convolutional neural network, which proves the modeling ability of the neural network method in the field of English speech rationality optimization recognition [27, 28]. However, the method for optimizing the rationality of English speech based on deep learning requires a large amount of training data to ensure its accuracy. Compared with the traditional method, it also requires a larger amount of calculation, which also limits its application in practice to a certain extent.

Under the guidance of the English speech rationality optimization recognition process, a network structure is built. Using the annotation data of the speech sentiment database and the text features obtained based on semantic analysis, a deep learning model is trained to realize the prediction of the optimal recognition of the speech sound from reading text to English speech. The pronunciation description is used as the final result output of the prediction model to guide the generation of subsequent acoustic parameters. The model comprehensively considers the influence of the context environment of different scale units at the chapter level, paragraph level, and sentence level, as well as the interaction between the various links in the process of optimizing the recognition of English sound rationality, forming a multilayer-nested composite network to support structural construction. First, an improved auxiliary traceless particle filtering algorithm is established based on the standard particle filtering algorithm, and the algorithm implementation steps are given. Then, the online parameter identification for the single-degree-of-freedom model is carried out and compared with the identification results of the traditional particle filter algorithm to verify the accuracy and calculation efficiency of the improved algorithm. Finally, the pseudostatic test of the seismic isolation support verifies the effectiveness of the improved auxiliary traceless particle filter algorithm for online identification of model parameters. The rest of this paper is organized as follows. Section 2 discusses optimal recognition modeling of English speech rationality, and optimized recognition model of English sound rationality based on improved particle filter algorithm is designed in Section 3. Experimental verification is discussed in Section 4. Section 5 concludes the paper with summary and future research directions.

## 2. Optimal Recognition Modeling of English Speech Rationality

At present, English speech rationality optimization recognition technology is facing huge development opportunities. First, with the development of devices with high computing power, more complex algorithms and models become possible; second, with the help of massive data on the Internet, it becomes easier to have corpus resources in real scenes, making the trained models more reasonable; finally, the rise and popularity of applications such as smart homes, car systems, and mobile devices make voice, a convenient human-computer interaction, more important. The optimal recognition of English speech rationality can be divided into four main parts, namely, data preprocessing and feature extraction, language model, acoustic model, and decoder, as shown in Figure 1.

The data preprocessing part is performed before the identification and decoding, including the signal pre-emphasis, framing, windowing, and other operations [29]. Research has found that the energy of speech is mainly in the low frequency range, which may lead to an excessive signal-to-noise ratio in the high frequency range. Preemphasis can enhance the high frequency band of speech to make its characteristics more prominent. English speech is a non-stationary time-varying signal, but it has a short-term stationary characteristic when the duration is short, so its signal processing needs to be performed after framing. By framing the signal, each frame can be regarded as a smooth signal. Generally speaking, the duration of each frame is between 10 and 30 ms and there will be overlap between frames. After framing, it will cause discontinuity at the beginning and end of each frame. Windowing is to highlight the signal in the middle and make it continuous. The window function usually chooses the Hamming window.

Feature extraction is mainly to extract some information that can reflect the commonality from the signal. When different speakers speak the same paragraph of text, the pronunciation will be different due to the difference in vocal organs and speaking habits. Feature extraction is to remove these characteristics. The feature used by the system in this paper is the MFCC feature, which is one of the commonly used features in the field of speech recognition.

The purpose of English speech rationality optimization recognition is to convert speech into text. After the system receives a piece of audio, it can find the most reasonable sequence of words to represent the information contained in this piece of audio. We define the speech signal as  $T$  and the text sequence as  $M$ ; then the optimization and recognition of English sound rationality need to be solved:

$$M = \arg \max P(M | T). \quad (1)$$

That is, on the premise of knowing the voice  $T$ , find the most likely text sequence  $M$ . According to the Bayesian formula, we can change the above formula to

$$M = \arg \max \frac{P(T | M)P(M)}{P(T)}, \quad (2)$$

For a particular piece of audio  $T$ ,  $P$  is fixed and therefore does not need to be considered in the optimization process. The above formula is the core formula of speech recognition, which can be seen as a combination of two parts, the language model  $P(M)$  and the acoustic model  $P(T | M)$ ; the language model represents a certain sequence of text  $M$  in language habits rationality.

The main function of the decoder is to use the trained language model and acoustic model to build a decoding network, search in the network, and finally find the optimal path that can explain the input speech and give the recognition result. The decoder can search through a given input feature sequence and finally find and use the decoding algorithm to best interpret the input audio text sequence for the search composed of the language model and the acoustic model.

Feature selection is a link that has a significant impact on the optimization of the recognition efficiency of English

speech. Deciding which acoustic features to use will largely affect the recognition rate of the system. The acoustic features commonly used in the optimization and recognition of English speech rationality include perceptual linear prediction, linear prediction coefficients, and Mel-frequency cepstral coefficients, and cepstral features are commonly used Mel-frequency cepstral coefficients. The Mel-frequency is the cepstral parameter extracted from the frequency domain of the scale. It has a high degree of similarity to the human auditory characteristics. Mel-frequency cepstral coefficient can perfectly combine the frequency selection characteristics of ear-to-sound and speech signal processing technology. It has strong robustness and noise resistance. It is one of the most effective features in the field of English speech sound optimization recognition.

The extraction process of English speech rationality optimization recognition features is shown in Figure 2.

Before extracting the English speech rationality optimization recognition feature for the audio signal, we need to perform relevant preprocessing operations on the data. The power of the voice signal will be very small when the frequency is large. The main energy is distributed in the low frequency band. This may cause the signal-to-noise ratio of the high frequency band to be too large. Pre-emphasis can enhance the high-frequency signal of the voice. Speech is a nonstationary signal, but it is stationary for a short period of time. The framing and windowing operations are to take advantage of this short-term stationarity to divide the long-term nonstationary signal into multiple frames of shorter stationary signals. Generally speaking, the length of each frame of speech is between 10 and 30 ms, and there is overlap between the frames, which can ensure the continuity of the signal. The window function used in this article is the Hamming window, which can be expressed as

$$M(n) = \begin{cases} 0.5 - 0.48 \cos \frac{2\pi n}{M-1}, & 0 \leq n \leq M \\ 0, & \text{OTHER.} \end{cases} \quad (3)$$

After preprocessing the English speech signal, we use fast Fourier transform to transform the obtained short-term signal  $p(1)$  into time frequency and calculate the short-term energy spectrum  $h(1)$ . Then we use the filter  $c(1)$  to process the short-term energy spectrum, followed by using the filter to process the short-term energy spectrum; the formula can be expressed as

$$M(j) = \begin{cases} \frac{j - p(1)}{c(1) - p(1)}, \\ \frac{h(1) - j}{j(1) - c(1)}. \end{cases} \quad (4)$$

Then, we make logarithmic changes to the output of the filter bank and finally through the discrete cosine transform to obtain the English speech rationality optimization recognition feature coefficients.

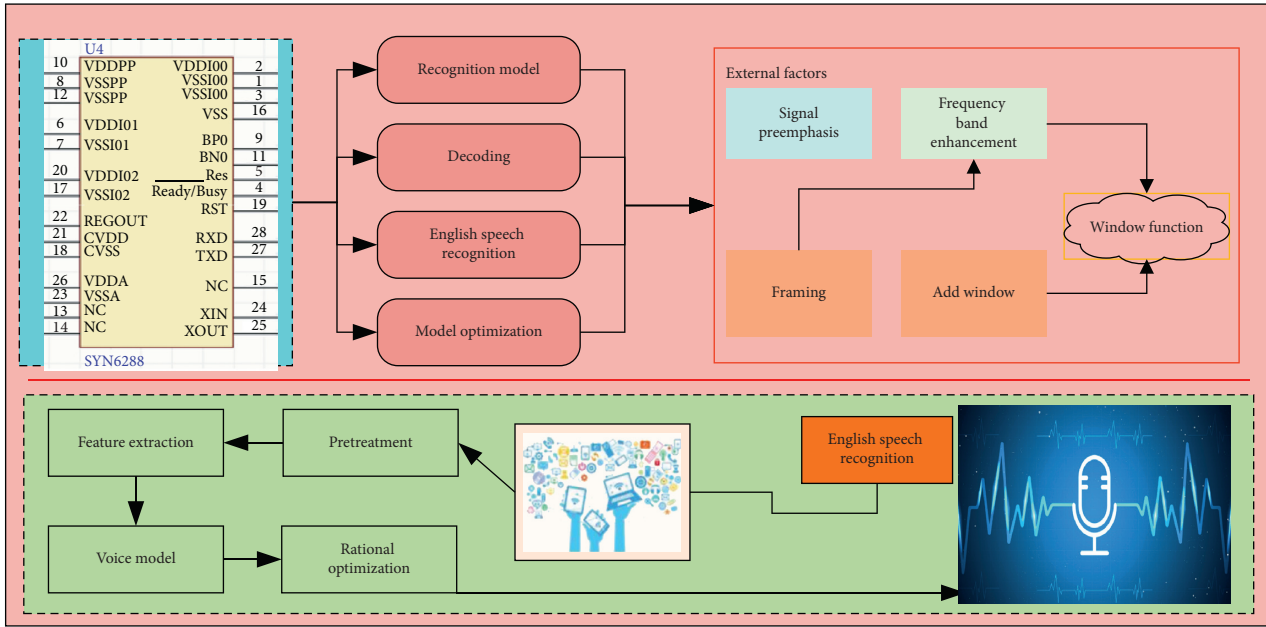


FIGURE 1: English speech rationality optimization recognition framework.

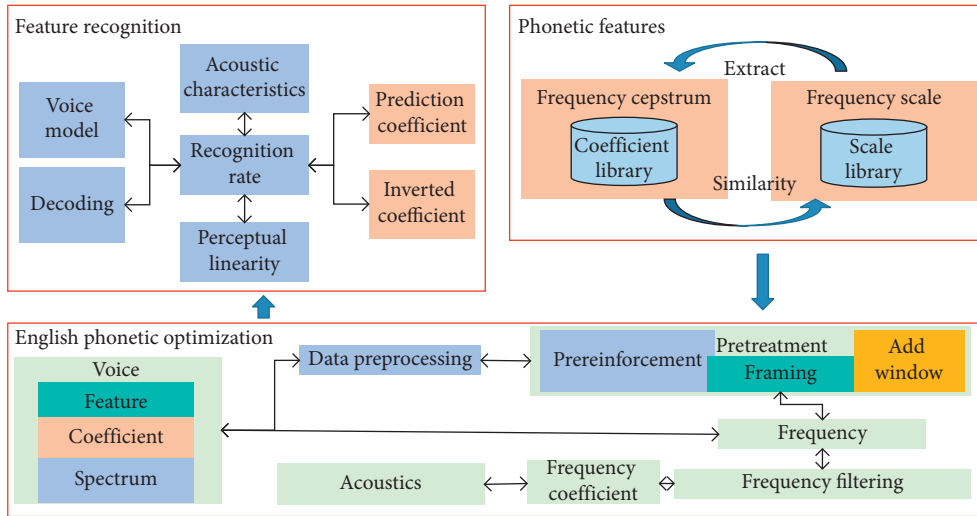


FIGURE 2: English speech rationality optimization recognition feature extraction process.

### 3. Optimized Recognition Model of English Sound Rationality Based on Improved Particle Filter Algorithm

3.1. Reasonable Optimization of Identification Parameters. The improved particle filter algorithm inherits the principle of the particle filter algorithm and is a complete nonlinear estimator that can identify arbitrary nonlinear model parameters. The improved particle filter algorithm model is a representative strong nonlinear model used to simulate the restoring force characteristics of structures and members. In this paper, the single-degree-of-freedom improved particle filter algorithm model is taken as the object, and a specific implementation method of applying the algorithm to online identification of nonlinear model

parameters is given to verify the algorithm recognition accuracy.

Combined with structural motion equation, particle filter algorithm model was improved.

Set the model parameter of the improved particle filter algorithm to the actual value of  $k_0 = 40 \text{ kN/m}$ ,  $n = 1.1 \text{ t}$ ; load the model with displacement control, input displacement excitation, and select the ground motion displacement record measured by the seismic station; the peak displacement is adjusted to 10 cm, as shown in Figure 3, vertical. The coordinates are displacement. Using the fourth-order numerical integration method to calculate the resilience of the improved particle filter algorithm model system, the integration step length is 0.01 s, and the integration number of steps is 4000 steps.

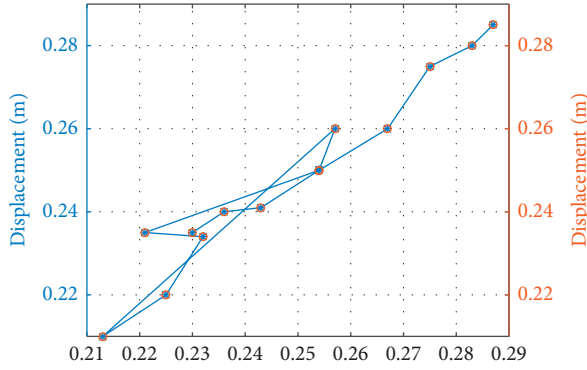


FIGURE 3: Displacement loading time history curve.

$z$  is the process noise;  $k$  is the actual loading speed;  $n$  is the number of times; the system state equation is

$$y = f(y, y') = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} - \begin{bmatrix} z \\ k \\ \dots \\ n \end{bmatrix}. \quad (5)$$

**3.2. Stepped Resampling Improvements.** In order to solve the problems of sample degradation and reduced estimation accuracy caused by resampling, this paper proposes a particle selection scheme based on the bias-corrected exponential weighted average algorithm and then effectively uses the remaining large-weight particles to complete the replication and new number addition ensure the final particles meets the initial total.

Aiming at the problem that standard resampling directly removes small particles and reduces the filtering performance, the following particle screening strategy is designed: (1) Calculate the weight of particles and arrange them in ascending order. (2) Use the EWA optimization algorithm based on deviation correction to calculate the sorted particle set. The average distribution curve of the expression is:

$$\gamma_i = \frac{\varepsilon \gamma_{i-1} + (1 - \varepsilon) \omega_t^i}{1 - \varepsilon^i}. \quad (6)$$

In the previous equation,  $\gamma_i, \gamma_{i-1}$  is the average weight of the first  $i, i-1$  particles,  $\varepsilon$  is the overshoot parameter,  $\omega_t^i$  is the weight of the  $i$ -th particle, and  $1 - \varepsilon^i$  is the deviation correction term. (3) Calculate the average value of the  $\gamma_i$  curve and judge the particle weight  $\gamma$  with the size of  $\omega_t^i$ ; if  $\omega_t^i \geq \gamma$ , the particle is left and then is discarded. The exponentially weighted average optimization algorithm takes into account the fact that large-weight particles are the decisive factor that affects the estimation accuracy, and the introduction of deviation correction items reduces the previous calculation errors, improves the accuracy of particle screening, and also ensures the effectiveness of replication and new particles.

The process of optimizing the recognition of English speech rationality is divided into two parts: the training stage and the formal stage. The training stage first selects 10

articles for independent labeling, and then several labelers compare the results and discuss the labeling rules. The three annotation results are summarized and integrated into the final annotation result of each document by the introduction method in the next section. The agreement rate of the three labeling results in different scales is shown in Table 1. It can be seen from the table that the labeling agreement rates of the three scales are similar, and the chapter-level agreement rate is slightly higher than the other two levels; the agreement rate between two people is better than that of the three people, but the common agreement rate is also close to 80%, reaching a usable level.

**3.3. English Speech Rationality Optimization Keyword Recognition Model.** Let the English speech rationality optimization keyword recognition system have the rejection function; a feasible method is to use the junk speech model to improve the decoding network. In the training phase, each phoneme model is trained with the corpus of the corresponding phoneme, and the junk speech model uses all corpus for training, so it can be considered that it represents the human voice model rather than a specific phoneme. All voices have a certain matching ability.

Real time is an important feature of human-machine voice interaction. Whether the system can recognize the voice in time will play a vital role in the promotion of user experience and applications. The main time-consuming of the keyword speech recognition system is the decoding efficiency of English speech, so how to improve the efficiency of the decoding algorithm is the key to improving the performance of the system. The Viterbi decoding algorithm needs to search all possible states of the entire decoding network every frame. The token-passing decoding algorithm we mentioned is a specific implementation of the Viterbi decoding algorithm, which has the characteristics of simple search process implementation, thereby reducing the decoding time and improving system performance. However, the token transfer algorithm also cannot avoid the shortcomings of the whole network search. All tokens of the decoding network in each frame need to be transferred between all possible states. When the number of system states increases, the decoding time will also be large, in addition to an increase in amplitude.

The token-passing decoding algorithm searches through the entire network to ensure that the best state path is found. After obtaining the voice characteristics of each frame, all tokens in the decoding network will jump, although this can guarantee that each path can be searched, but it also limits the performance of the algorithm. Performing a network-wide search will consume a lot of computing resources on some paths with very low probability. In fact, these paths have a long distance from the token value at the beginning due to their low matching with the speech to be recognized. Lower than other tokens, the probability of these paths gaining an advantage in subsequent competition is very low, so these tokens can be discarded earlier to prevent their passing down, thereby reducing the consumption of computing resources and improving algorithm performance.

TABLE 1: Statistics of the coincidence rate of the labeling results.

|                 | A&B    | A&C    | B&C    | A&B&C  |
|-----------------|--------|--------|--------|--------|
| Chapter level   | 0.8264 | 0.843  | 0.8525 | 0.7821 |
| Paragraph level | 0.8275 | 0.8392 | 0.8483 | 0.7798 |
| Sentence level  | 0.8267 | 0.8378 | 0.8468 | 0.7788 |
| All             | 0.8269 | 0.8389 | 0.848  | 0.7795 |

“A,” “B,” and “C” represent the three labeling personnel.

In the specific implementation, the maximum number of tokens should be set reasonably according to the complexity of the network. If the maximum number of tokens is set too large, the system performance will cause many paths to be discarded per frame. The improvement is very limited. However, the maximum number of tokens is set too small to include the best path, which leads to a reduction in the accuracy of system identification. Therefore, when setting the maximum number of tokens, the performance and accuracy should be taken into consideration.

#### 4. Experimental Verification

In order to verify that the algorithm in this paper has good filtering performance, MATLAB simulation software is used to compare and analyze the number of particles with 50, 100, and 150 based on the univariate dynamic change filter model. Based on the improved particle filter algorithm, unified particle filter algorithm, scheduling particle filter algorithm, and particle filter algorithm, the English speech rationality optimization keyword recognition model’s online parameter recognition results obtained by four algorithms are compared, as shown in Figure 4. It can be seen that the four algorithms for  $k_0$  and  $n_2$  parameters have the same recognition effect, basically converging to the true value, and the convergence speeds are roughly similar. Among them, the recognition value obtained by the improved particle filter algorithm is the best. The improved particle filter algorithm has higher nonlinear transformation accuracy in importance sampling, enriches the diversity of particles during the resampling process, and effectively weakens the degradation of particles. The performance of the algorithm determines the pros and cons of the recognition result of the calculation example, which is of universal significance. Therefore, when the actual value is of the keyword recognition model parameters for English sound rationality optimization in the calculation example changes, under the same conditions, the recognition results of the four algorithms will still have similar rules.

The particle filter algorithm and its improved scheduling particle filter algorithm and the unified particle filter algorithm and the improved particle filter algorithm are essentially random parameter identification algorithms. The four algorithms are all based on the Monte Carlo random sampling method. Therefore, even under the same condition of initial parameter values, the parameter identification values obtained by each algorithm in each simulation are different; that is, the parameter identification results are random. In order to test the randomness of the recognition results of the algorithm, four independent filtering

algorithms were used for 10 independent simulations. The statistical recognition results compared and analyzed the recognition accuracy and convergence of different algorithms, which was more convincing. In the 10 independent simulations in this example, the initial parameters of the system English sound rationality optimization recognition model and the initial parameters of the algorithm are the same. The randomness mainly comes from the randomness of the particles produced by the algorithm. The relationship between the root mean square error of the four algorithms’ parameter identification values and the number of simulations is shown in Figure 5. The abscissa in the figure is the number of simulations.

It can be seen from Figure 5 that the overall error of the parameter identification of the improved particle filter algorithm is significantly lower than that of the particle filter algorithm, the scheduling particle filter algorithm, and the unified particle filter algorithm, and the error fluctuation range is significantly reduced. It can be seen that the improved particle filtering algorithm uses the latest observation information to modify the particles and at the same time increases the particle diversity by introducing auxiliary factors. Therefore, the improved particle filter algorithm is significantly higher than the recognition accuracy of the particle filter algorithm, scheduling particle filter algorithm, and unified particle filter algorithm.

As shown in Figure 6, the root mean square error mean and relative error mean of the online independent parameter identification value was calculated under independent simulations. It can be clearly seen that the mean value of the root mean square error and the relative error of the parameter identification value of the improved particle filtering algorithm proposed in this paper are smaller than those of the particle filtering algorithm as a whole. The root mean square error of the four sets of parameter identification values obtained by the improved particle filter algorithm is overall lower than those of the particle filter algorithm, scheduling particle filter algorithm, and unified particle filter algorithm. The error is reduced by 76%, 38%, and 23%, respectively, and, in the improved particle filter algorithm, the relative errors of the obtained four sets of parameter identification values as a whole are reduced by 22%, 16%, and 14% compared to the particle filter algorithm, scheduling particle filter algorithm, and unified particle filter algorithm. It can be seen that the average root mean square error and the average relative error of the four sets of parameter identification values in 10 simulations indicate that the accuracy of the improved particle filtering algorithm is higher than that of the other three algorithms. It should be noted that the algorithm needs to pay more time for calculation while achieving higher parameter recognition accuracy.

When we only use the improved English speech rationality of the particle filtering algorithm to optimize the keyword recognition modeling model for token-passing decoding, because there is no adjustable parameter, the equal error rate cannot be measured, and the false rejection rate of the system measured by the test set is only 1.24%, but the false alarm rate is as high as 76.05%, which shows that the system with only the improved particle filter algorithm

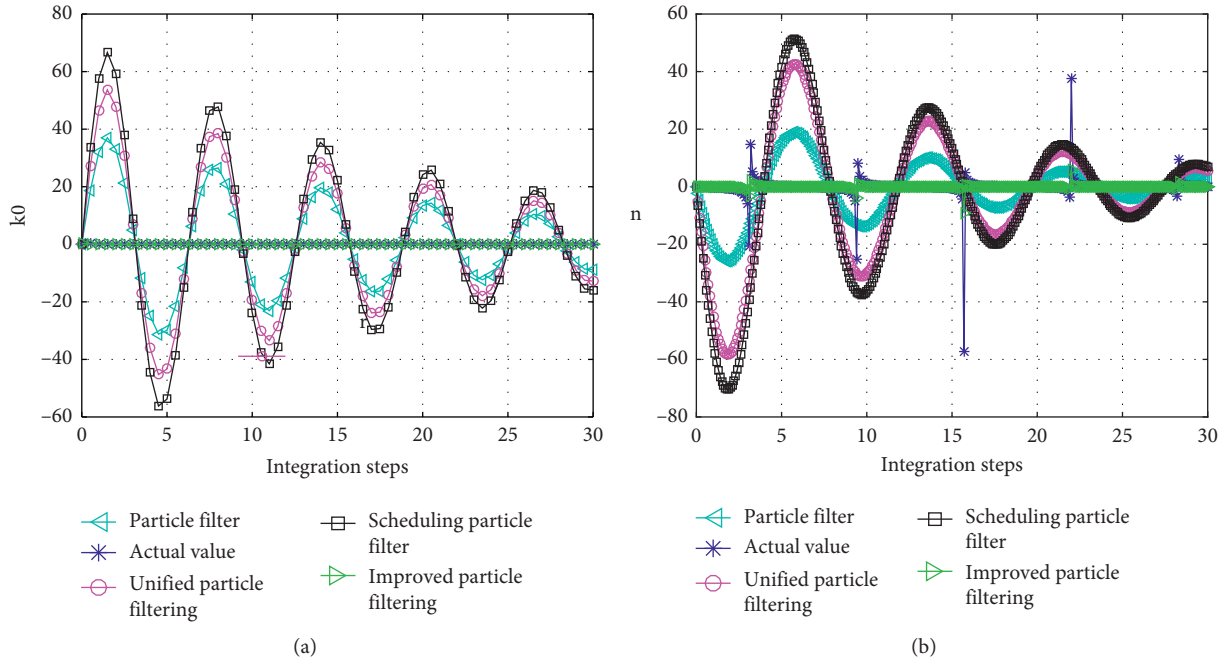


FIGURE 4: English speech rationality optimization keyword recognition value. (a)  $k_0$ . (b)  $n$ .

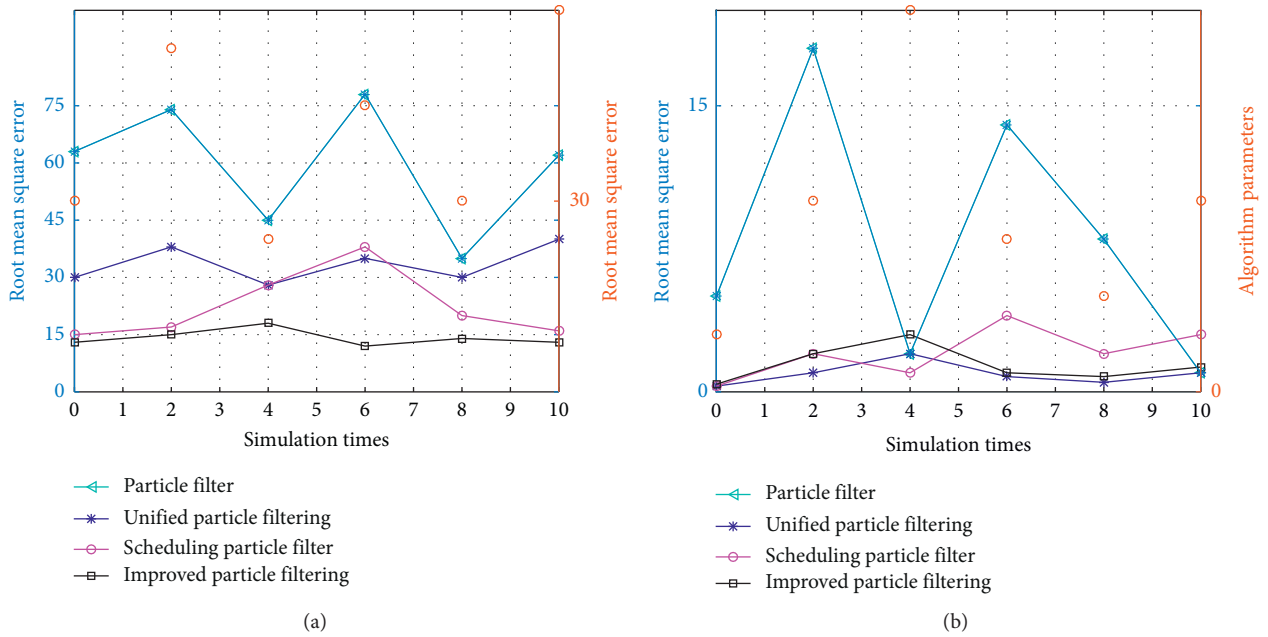


FIGURE 5: Root mean square error of model parameter identification value. (a)  $k_0$ . (b)  $n$ .

model is very weak for nonkeyword speech recognition, and it is difficult to apply it in practice. We use the verification set to adjust the parameters and use the test set to test the results and finally measured the equal error rate of several different systems. The results are shown in Table 2.

In this paper, the improved particle filtering algorithm proposed in this paper optimizes the stability of English speech keyword optimization modeling, which indicates whether the optimal token on the decoding path can maintain a stable advantage in the competition during the

decoding process. If the decoding path is the correct keyword path, then it should be able to stay ahead of the competition by constantly accumulating advantages. If the decoding path is wrong, the optimal token on this path will not be compared to other paths, and there is a big difference.

As shown in Figure 7, when the voice to be recognized is a keyword, the optimal token on the decoding path of each frame can stably maintain an advantage in the competition, and this advantage will increase as the number of decoded frames increases.

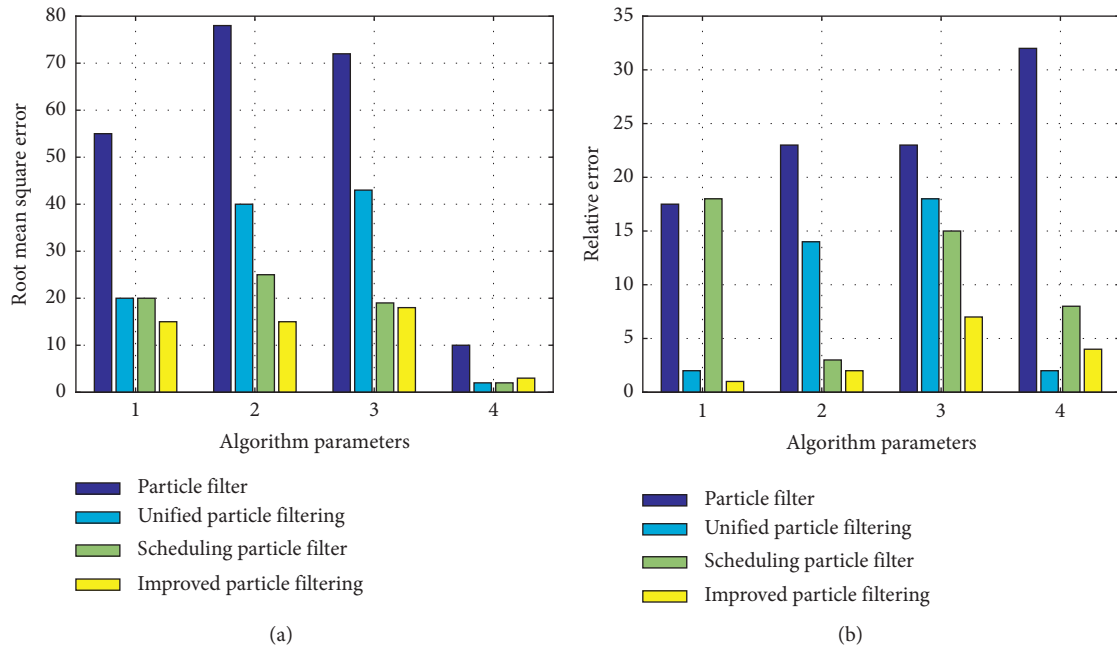


FIGURE 6: Root mean square error of parameter identification value and mean value of relative error. (a) Root mean square error. (b) Relative error.

TABLE 2: Different English pronunciation rationality optimization recognition error rate.

| Method                                    | Error rate (%) |
|---|----------------|
| Particle filtering algorithm [4]          | 16.84          |
| Scheduling particle filter algorithm [12] | 12.83          |
| Unified particle filtering algorithm [15] | 7.83           |
| Improved particle filtering algorithm     | 5.73           |

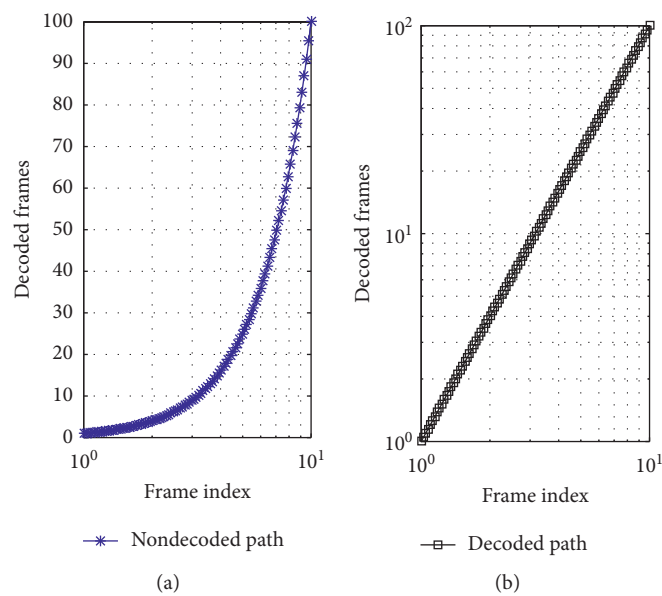


FIGURE 7: Stability data when the voice to be recognized is a keyword.



In this paper, the visibility modeling strategy of the deep neural network in the middle layer can effectively integrate prior knowledge into the network structure, thereby improving the performance of the network; optimization measures such as feature dimension balance adjustment and sample balance adjustment can solve the network learning process to a certain extent. The fusion problem between different types of features solve sample sparseness

## 5. Conclusion

In this paper, by introducing an improved particle filtering algorithm, the English speech rationality optimization recognition model enables the system to reject nonkeyword speech, which can be more widely used in actual scenarios. In order to improve the recognition accuracy and rejection ability of the speech keyword recognition system and make the application of the system more mature, on the basis of studying the principle of decoding algorithm, the data generated by the token transfer process in the preliminary experiment is analyzed, and finally two methods of consistency and stability were proposed and used to evaluate and correct the system identification results; and related experiments were set up to test the feasibility of the proposed method. The improved English speech rationality optimization recognition model with improved particle filter algorithm has high positioning accuracy and good stability, and its effectiveness is verified through experimental verification.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] C. Bai and C. Du, "Design and simulation of a collision-free path planning algorithm for mobile robots based on improved ant colony optimization," *Ingénierie des systèmes d'information*, vol. 24, no. 3, pp. 331–336, 2019.
- [2] M. Yuan, "Jitter buffer control algorithm and simulation based on network traffic prediction," *International Journal of Wireless Information Networks*, vol. 26, no. 3, pp. 133–142, 2019.
- [3] Y. Guo, L. Miao, and Y. Lin, "A novel EM implementation for initial alignment of SINS based on particle filter and particle swarm optimization," *Mathematical Problems in Engineering*, vol. 2019, Article ID 6793175, 12 pages, 2019.
- [4] J. Li, W. Lu, H. Wang et al., "Identification of groundwater contamination sources using a statistical algorithm based on an improved Kalman filter and simulation optimization," *Hydrogeology Journal*, vol. 27, no. 8, pp. 1–13, 2019.
- [5] K. Han and H. Zhang, "Improved particle filter based on fruit fly optimization and its application in target tracking," *Hunan Daxue Xuebao/Journal of Hunan University Natural Ences*, vol. 45, no. 10, pp. 130–138, 2018.
- [6] L. Peng, Y. Wu, and L. Huang, "Opto-electric target tracking algorithm based on local feature selection and particle filter optimization," *Concurrency and Computation: Practice and Experience*, vol. 30, no. 22, pp. e4670.1–e4670.11, 2018.
- [7] Y. Jiang, Z. He, and A. Liu, "A complementary optimal operation strategy of wind power-hydropower based on improved biogeography-based optimization algorithm," *Dianli Xitong Baohu Yu Kongzhi/power System Protection & Control*, vol. 46, no. 10, pp. 39–47, 2018.
- [8] B.-R. Ding, X.-H. Qu, Y.-L. Chen et al., "Application and research of mechanical design optimization based on genetic algorithm kinematics simulation technology," *Journal of Intelligent & Fuzzy Systems*, vol. 34, no. 2, pp. 871–878, 2018.
- [9] W. Niu and H. Cao, "Two-stage beam structure damage detection based on improved particle swarm optimization," *Tumu Jianzhu Yu Huanjing Gongcheng/journal of Civil Architectural & Environmental Engineering*, vol. 40, no. 6, pp. 123–130, 2018.
- [10] O. Wang and O. Yang, "Research on industrial assembly line balancing optimization based on genetic algorithm and witness simulation," *International Journal of Simulation Modelling*, vol. 16, no. 2, pp. 334–342, 2017.
- [11] Y. Yun and S. K. Kim, "Simulation of aseismic structure optimization of buildings based on finite element numerical simulation and improved genetic algorithm," *Boletin Tecnico/technical Bulletin*, vol. 55, no. 19, pp. 171–177, 2017.
- [12] S. I. Gangquan, L. I. Shuiwang, S. Jianquan et al., "Least squares support vector machine parameters optimization based on improved fruit fly optimization algorithm with applications," *Journal of Xi'an Jiaotong University*, vol. 51, no. 6, pp. 14–19, 2017.
- [13] G. Wang, R. Ma, Q. Meng et al., "Maximum non-gaussianity estimation revisit: uniqueness analysis from the perspective of constrained cost function optimization," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 32, no. 5, pp. 1851004.1–1851004.18, 2018.
- [14] X. Wang, Y. Ma, and Y. Yang, "Structural design optimization based on hybrid time-variant reliability measure under non-probabilistic convex uncertainties," *Applied Mathematical Modelling*, vol. 69, pp. 330–354, 2019.
- [15] J. Zeng and J. Li, "A Bayesian belief network approach for mapping water conservation ecosystem service optimization region," *Journal of Geographical Sciences*, vol. 29, no. 6, pp. 1021–1038, 2019.
- [16] S. Jingbo, Z. Yue, and L. Xiaoxiao, "Optimization of test resources for 3D SoC based on game theory," *Journal of China Universities of Posts & Telecommunications*, vol. 26, no. 1, pp. 69–78, 2019.
- [17] C. Tian, Q. Lu, and Z. Xiao, "Cutting parameter optimization for machining operations considering carbon emissions," *Journal of Cleaner Production*, vol. 208, no. 1658, pp. 937–950, 2019.
- [18] L. Jiang, Y. Wang, and C. Liu, "Hierarchical nondeterministic optimization of curvilinearly stiffened panel with multi-cutouts," *Aiaa Journal*, vol. 56, no. 10, pp. 4180–4194, 2018.
- [19] X. Zhu, Y. Ding, and J. Zhuang, "Multi-Objective optimization design and multi-physics analysis a double-stator permanent-magnet doubly salient machine," *Energies*, vol. 11, no. 8, pp. 2130–2138, 2018.
- [20] A.-L. Jin, W. Song, and W. Zhuang, "Auction-based resource allocation for sharing cloudlets in mobile cloud computing," *IEEE Transactions on Emerging Topics in Computing*, vol. 6, no. 1, pp. 45–57, 2018.
- [21] Y. Sun, R. Liu, and Y. Fan, "Analytical modeling and damping optimization for a thin plate partially covered with hard

- coating,” *Archive of Applied Mechanics*, vol. 88, no. 6, pp. 897–912, 2018.
- [22] Y. Lv, C. Li, and G. Zhang, “Data driven parallel prediction of building energy consumption using generative adversarial nets,” *Energy and Buildings*, vol. 186, no. 3, pp. 230–243, 2019.
- [23] R. Pan, Y. Zhang, J. Ding et al., “Rationality optimization of tool path spacing based on dwell time calculation algorithm [J],” *International Journal of Advanced Manufacturing Technology*, vol. 84, no. 9-12, pp. 2055–2065, 2016.
- [24] Q. Liu, G. Zhou, and F. Guo, “Simulation-optimization method based on rationality evaluation for waste load allocation in Daliao river,” *Environmental Earth Sciences*, vol. 73, no. 9, pp. 5193–5209, 2015.
- [25] D. Liu, W. Bu, and D. Xu, “Multi-objective shape optimization of a plate-fin heat exchanger using CFD and multi-objective genetic algorithm,” *International Journal of Heat and Mass Transfer*, vol. 111, no. 4, pp. 65–82, 2017.
- [26] Z. Feng, G. Wang, and Z. Feng, “Optimization analysis of primary mirror in large aperture telescope based on workbench,” *International Journal of Climatology*, vol. 34, no. 12, pp. 3417–3425, 2015.
- [27] R. Zhou, S. He, and R. Song, “Optimization for service routes of pallet service center based on the pallet pool mode,” *Computational Intelligence and Neuroscience*, vol. 2016, no. 3, 11 pages, Article ID 5691735, 2016.
- [28] T. Q. Bao, B. S. Mordukhovich, and A. Soubeyran, “Fixed points and variational principles with applications to capability theory of wellbeing via variational rationality,” *Set-Valued and Variational Analysis*, vol. 23, no. 2, pp. 375–398, 2015.
- [29] S. Liu, “Multi-objective optimization design method for the machine tool’s structural parts based on computer-aided engineering,” *The International Journal of Advanced Manufacturing Technology*, vol. 78, no. 5-8, pp. 1053–1065, 2015.

## Research Article

# Numerical Simulation of Ambiguity Resolution in Multiple Information Streams Based on Network Machine Translation

Lei Wang<sup>1</sup> and Qun Ai <sup>2</sup>

<sup>1</sup>Department of Foreign Language, Jilin Business and Technology College, Jilin, Changchun 130000, China

<sup>2</sup>Department of Basic Education, Jilin University, Jilin, Changchun 130000, China

Correspondence should be addressed to Qun Ai; [sabirin9917@mails.jlu.edu.cn](mailto:sabirin9917@mails.jlu.edu.cn)

Received 6 April 2020; Revised 22 July 2020; Accepted 3 August 2020; Published 17 August 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Lei Wang and Qun Ai. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In natural language, the phenomenon of polysemy is widespread, which makes it very difficult for machines to process natural language. Word sense disambiguation is a key issue in the field of natural language processing. This paper introduces the more common statistical learning methods used in the field of word sense disambiguation. Using the naive Bayesian machine learning method and the feature vector set extracted and constructed by the Dice coefficient method, a semantic word disambiguation model based on semantics is realized. The results of comparative experiments show that the proposed method is better compared with known systems. This paper proposes a method for disambiguation of word segmentation in professional fields based on unsupervised learning. This method does not rely on professional domain knowledge and training corpus and only uses the frequency, mutual information, and boundary entropy information of the string in the test corpus to solve the problem of word segmentation ambiguity. The experimental results show that these three evaluation standards can solve the problem of word segmentation ambiguity in professional fields and improve the effect of word segmentation. Among them, the segmentation result using mutual information is the best, and the performance is stable.

## 1. Introduction

Word sense disambiguation is to determine the most exact semantics of a polymorphic word according to its context and locale [1]. Chinese is the most spoken language in the world, and it is difficult to avoid the problem of polysemy. According to statistics, ambiguous words account for about 14% of the total vocabulary of the Chinese dictionary, and these ambiguous words are often commonly used words [2]. Statistics on the authoritative Chinese disambiguation corpus show that these ambiguous words are used very frequently, about 42% [3]. The universality of ambiguous word distribution makes word sense disambiguation an important link in many applications related to natural language processing, such as machine translation, information extraction, and content analysis [4, 5].

The earliest Chinese word segmentation method was a word segmentation method based on “lookup dictionary”

[6]. The idea of this method is to read the entire Chinese sentence and then mark all the words in the dictionary separately. When encountering a compound word (such as Peking University), the longest word match is found. The string is split into individual words. This word segmentation method is not efficient, but its proposal lays the foundation for Chinese automatic word segmentation technology [7]. Relevant scholars have theorized the Chinese word segmentation method and proposed the “minimum number of words” segmentation theory; that is, each sentence should be segmented with the least number of words [8]. This word segmentation method is an improvement on the “word dictionary” word segmentation method, which has promoted the development of Chinese word segmentation technology. Researchers believe that, for computers to reach the level of natural language processing by humans, computers must be able to automatically disambiguate ambiguous words in a specific context and choose the most precise

meaning [9–11]. Although the word sense disambiguation system was only a subsystem in the machine translation system at that time, the context window and semantic consistency proposed by it were still the basis for the current research on word sense disambiguation. Among them, the context in which ambiguous words are located is an important condition that affects the accuracy of word sense disambiguation [12]. After that, with increasing attention to word sense disambiguation, more and more experts and scholars have proposed more solutions to word sense disambiguation [13, 14]. With the deepening of research, due to the lack of relevant resources and calculation methods at that time, scholars realized that word sense disambiguation was a very complicated problem and it was difficult to overcome [15]. Bar-Hillel pointed out that word sense disambiguation was impossible to achieve with the technical conditions at the time and was theoretically not feasible [16]. The method of automatically expanding WordNet uses a lot of semantic relationships from Wikipedia resources to expand WordNet. First, it establishes a mapping between Wikipedia pages and WordNet and then converts the pages into WordNet. Test results show that this method improves baseline and can use more disambiguation information to achieve higher disambiguation goals [17]. With the continuous updating of technology and continuous improvement of machine performance, techniques such as machine learning and corpus are used in word sense disambiguation. During this period, supervised, semisupervised, and unsupervised word sense disambiguation techniques have been fully developed. Relevant scholars have proposed a completely unsupervised method based on the topic word frequency estimation model [18]. This method can be applied to any part of the speech manuscript without the need for a hierarchical corpus or parallel text. The library is highly portable. Furthermore, the effectiveness of the method on the main semantic learning and semantic distribution acquisition tasks is proved. Relevant scholars have studied a new and effective fuzzy classification system and applied this system to word sense disambiguation [19, 20]. The system iteratively adjusts the weights of fuzzy rules and adjusts the classifier by adding weights to the rules [21]. Compared with other classification systems, this classification system has achieved good results. The advantage of the unsupervised word segmentation method is that it does not rely on dictionaries and training corpus and does not require training. It can be used for unregistered word discovery without word formation rules. The disadvantage is that this kind of method cannot find low-frequency words, the upper limit of the word segmentation effect is about 0.85, and the word segmentation effect cannot meet the practical application. In practical applications, the unsupervised word segmentation method is generally not used alone. This type of word segmentation method can be used for common word discovery, new words discovery, and other issues and assist in improving the word segmentation effect of existing dictionaries and training corpus word segmentation methods.

For word sense disambiguation, these classification models cannot be simply used, and corresponding improvements should be made on the basis of the original

models. This paper uses the sliding word window to extract the semantic-related features of words in the word window and constructs a Bayesian word sense disambiguation model based on semantics to perform the word sense disambiguation experiment. The Bayesian word sense disambiguation classifier based on semantic information is constructed using three different vector sets in feature extraction. The new method is used to verify the word sense disambiguation performance. This paper proposes to use unsupervised learning to solve the problem of word segmentation ambiguity in professional fields. The frequency of the strings in the test corpus, the mutual information of the strings, the boundary entropy of the strings, and the boundary entropy of the single words are used as the evaluation criteria to resolve the ambiguity problem. Experiments show that these three evaluation standards can solve the problem of word segmentation ambiguity in professional fields to varying degrees. The rest of this paper is organized as follows. Section 2 studies the semantic-based disambiguation model of machine translation. Section 3 analyzes the method of disambiguation of word segmentation in unsupervised professional fields. Section 4 summarizes the full text and points out future research directions.

## 2. Semantic-Based Disambiguation Model for Machine Translation

### 2.1. Machine Translation System

*2.1.1. Language Model.* The language model treats sentences as strings, and the probability that each word in the sentence appears as a character in the sentence is random. For a given string  $w_1, \dots, w_n$ , the probability of its occurrence can be expressed as

$$P(W) = \prod_{k=0}^{n-1} P(w_k | w_1, w_2, \dots, w_{k-1}). \quad (1)$$

$N$ -grams are generally used in language models to calculate the probability of an entire string. The probability that the  $N$ th word appears in the  $N$ -gram model is only related to the first  $N-2$  words; that is,

$$P(w_n | w_1^{n-1}) = P(w_n | w_{n-N+2}^{n-2}). \quad (2)$$

Linguistic studies have shown that the emergence of the current word is strongly dependent on many of the words before it [22]. Language models provide a way to calculate the probability of a string appearing. The disadvantage is that a large-scale corpus is needed to determine the parameters of the model.

*2.1.2. Translation Model.* In order to construct a translation model for a phrase machine translation system, we first need to calculate the phrase translation probability and dictionary probability. The phrase translation probability indicates the probability that a phrase on the source language side is translated into a phrase on the target language side. To obtain these two probabilities, four operations need to be performed on the parallel aligned bilingual corpus: word

alignment, word score, phrase extraction, and phrase score. A schematic diagram of a network bilingual information processing system for machine translation is shown in Figure 1.

According to the probability score calculation of the translated sentence, the longer the sentence, the smaller the probability. So when translating, decoders for phrase machine translation tend to choose shorter sentences. Therefore, longer target sentences need to be compensated. The length penalty model calculates the number of words in the translation as a penalty value and adds it to the model, which can be expressed as

$$\Pr(e) = \exp(I). \quad (3)$$

In the formula,  $I$  represents the number of translated words. The word punishment model can choose the length of the translation.

**2.2. Word Sense Disambiguation Method Based on Statistical Learning.** The Bayesian method is implemented using probability calculations. It is inferred from the probability that an event has happened in the past. The Bayesian method is applied to the word sense disambiguation problem as follows:

$$P(S_i | \text{Context}) = \frac{P(S_i)P(\text{Context} | S_i)}{\sum_{j=1}^m P(S_j)P(\text{Context} | S_j)}, \quad i = 1, 2, 3, \dots, m. \quad (4)$$

Word sense disambiguation based on Bayesian method is to judge the classification of word meaning based on the size of the posterior probability. Among them, Context is the context in which the ambiguous word  $w$  is located. Context is composed of word units on both sides of  $w$ , which provide necessary guidance information for the disambiguation process. The ambiguous word  $w$  has  $m$  semantic categories  $S_1, S_2, \dots, S_m$ . In Context, its true semantic category is  $S_j$ .  $P(X)$  is the probability that  $X$  will occur. In the Context, if the probability of the ambiguous word  $w$  taking the semantic category  $S_j$  is greater than any other semantic category  $S_i$  ( $j \neq i$ ), then the semantic category of the word  $w$  should be determined as  $S_j$ . The Bayesian method has the following two commonly used classifications in practical applications.

**2.2.1. Multivariate Bayes Based on Berle Effort.** First, we find the feature vector  $F = (t_1, \dots, t_m)$  through feature selection, where  $t_i$  is the label of each feature. For the dichotomy, this method deduplicates the corpus sample  $d$  to obtain the label set  $X = (x_1, \dots, x_m)$ , where  $x_i$  is the categorical variable value, which is 0 or 1 for the dichotomous case. According to the results of  $m$  Bernoulli experiments, the disambiguation corpus belongs to the category  $C$ :

$$P(t | c) = \frac{(M_{t,c} + 1)}{(M_c + 2)}. \quad (5)$$

To prevent the denominator from being zero, we use Laplace's method for smoothing.  $M_{t,c}$  is the number of texts that belong to category  $c$  and feature  $t_i$ , and  $M_c$  is the number of texts that belong to category  $c$ .

In order to prevent the result from underflowing, logarithmic processing is generally applied to the probability value in the application. Word sense disambiguation classification criterion is  $T=0$ ; it judges an ambiguity word classification.

**2.2.2. Polynomial Bayes Based on Boolean Attributes.** Polynomial Bayesian method based on Boolean attribute is similar to polynomial Bayesian method based on word frequency. The attribute value in Boolean attribute is Boolean type. However, when the feature is  $x_i = 0$ , the feature term is not added to the calculation of the conditional probability. At this time,

$$P(\bar{X} | C) = \prod_{i=0}^{m-1} p(t_i | C)^{x_i}. \quad (6)$$

And the Laplacian smoothing method is different.

**2.3. Establishment of Word Sense Disambiguation Model.** Most machine learning methods have been applied to the field of word sense disambiguation, which can be divided into discriminative models and generative models according to different model learning methods. In terms of accuracy and efficiency of disambiguation, the Bayesian model has a good balance, and the good robustness of the Bayesian model is also a key point for many word sense disambiguation models to use the model. The English combination ambiguity resolution framework is shown in Figure 2.

According to the characteristics of Chinese word sense disambiguation, this paper uses the sliding word window method to open the word window according to the position of the target ambiguous word, uses the Dice coefficient method to obtain the semantic information of the feature words in the word window as the disambiguation feature, and constructs a feature vector set. The feature vector set is applied to the Bayesian model to obtain a semantic sense disambiguation model based on semantic knowledge.

In the language environment where the ambiguous word is located, the final interpretation of the ambiguous word is determined based on the maximization of the posterior probability, if

$$P(S_j | \text{Context}) \geq P(S_i | \text{Context}), \quad (i = 1, 2, \dots, m). \quad (7)$$

The semantics of the ambiguous word  $w$  is  $S_j$ . Among them, Context is the context in which the ambiguous word  $w$  is located. Generally, Context is composed of word units on both sides of  $w$ , which provide necessary guidance information for the disambiguation process. The ambiguous word  $w$  has  $m$  semantic categories  $S_1, S_2, \dots, S_m$ . In Context, its true semantic category is  $S_j$ .  $P(X)$  is the probability of  $X$  appearing. In the context, if the probability of the ambiguous

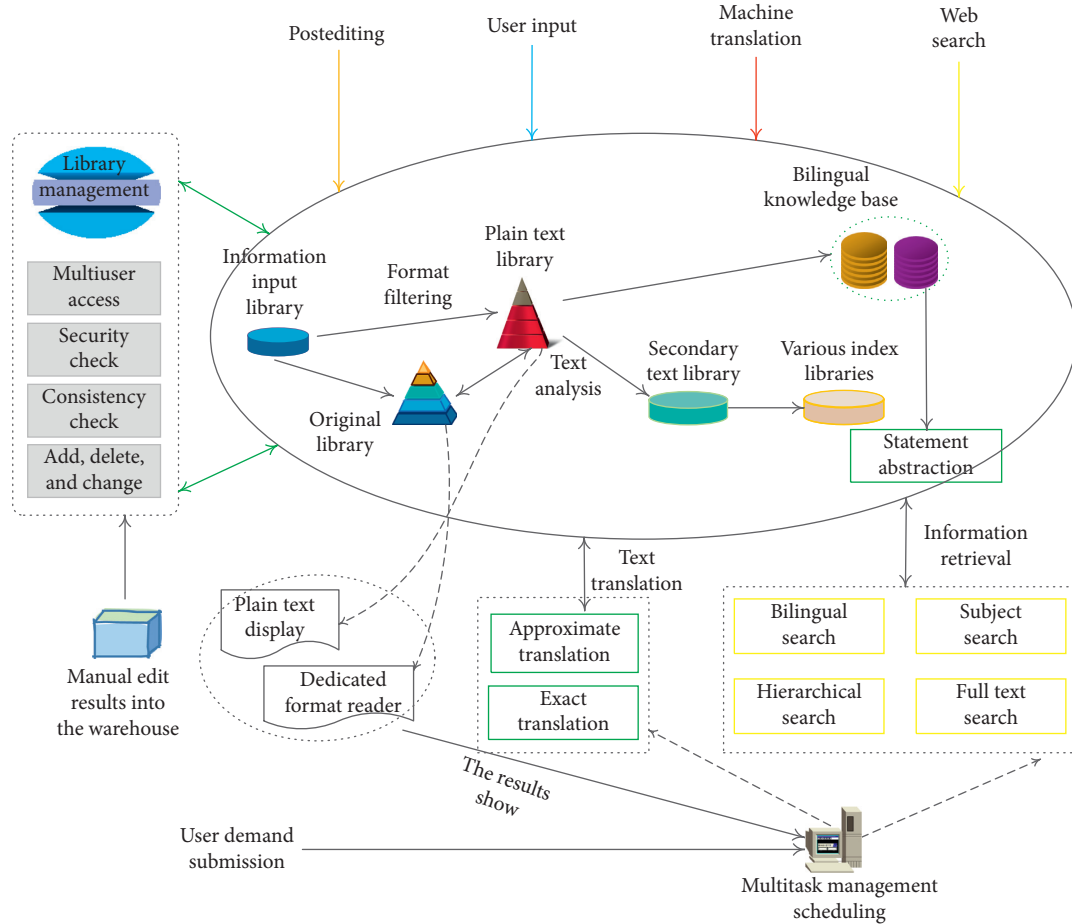


FIGURE 1: Network bilingual information processing system for machine translation.

word  $w$  taking the semantic category  $S_j$  is greater than any other semantic category  $S_i$  ( $j \neq i$ ), the semantic category of the word  $w$  should be determined as  $S_j$ . Bayesian decisions have the smallest probability of error. For the ambiguous word  $w$ , it has  $m$  semantics  $S_1, S_2, \dots, S_m$ , and the context disambiguation features are FL and FR, respectively. Among them, FL and FR are semantic class codes. The semantic code of "Synonym Cilin" is divided into three layers: FL = fl1fl2fl3; FR = fr1fr2fr3.

The process of word sense disambiguation based on semantic knowledge is as follows:

- (1) We analyze the corpus to obtain sentence information, segmentation information, and part-of-speech information.
- (2) We use the sliding word window method to find the position of the target ambiguous word and use it as the center to open the word window on both sides to obtain the word window segmentation.
- (3) We use "Synonyms Word Forest" as a semantic dictionary to obtain the semantic code set of word segmentation in the word window.
- (4) We use the Dice coefficient method to determine the word segmentation code. The feature vector set is constructed by three different feature extraction

methods: first-level semantic code, three-level semantic code, and morphological information.

- (5) We construct a word sense disambiguation model based on semantic knowledge by using the feature vector set extracted from the training corpus.
- (6) We disambiguate the test corpus by constructing a good word sense disambiguation model. Three different tests are performed on different feature vector sets.

**2.4. Construction and Analysis of Numerical Simulation Experiments.** We select 10 more polysemous words for comparison. These words include dichotomous vocabulary and multiple vocabularies. We made statistics on their various data, and the data situation is shown in Figure 3.

We obtain the experimental corpus and preprocess the corpus. The preprocessing part mainly includes analyzing the corpus and locating the target ambiguous words. The sliding word window method is used to obtain the lexical information set near the target ambiguous words. The feature processing part mainly includes acquiring the semantic knowledge set and determining the semantic knowledge by the Dice coefficient method. Semantic codes can be divided into three categories: large, medium, and small.

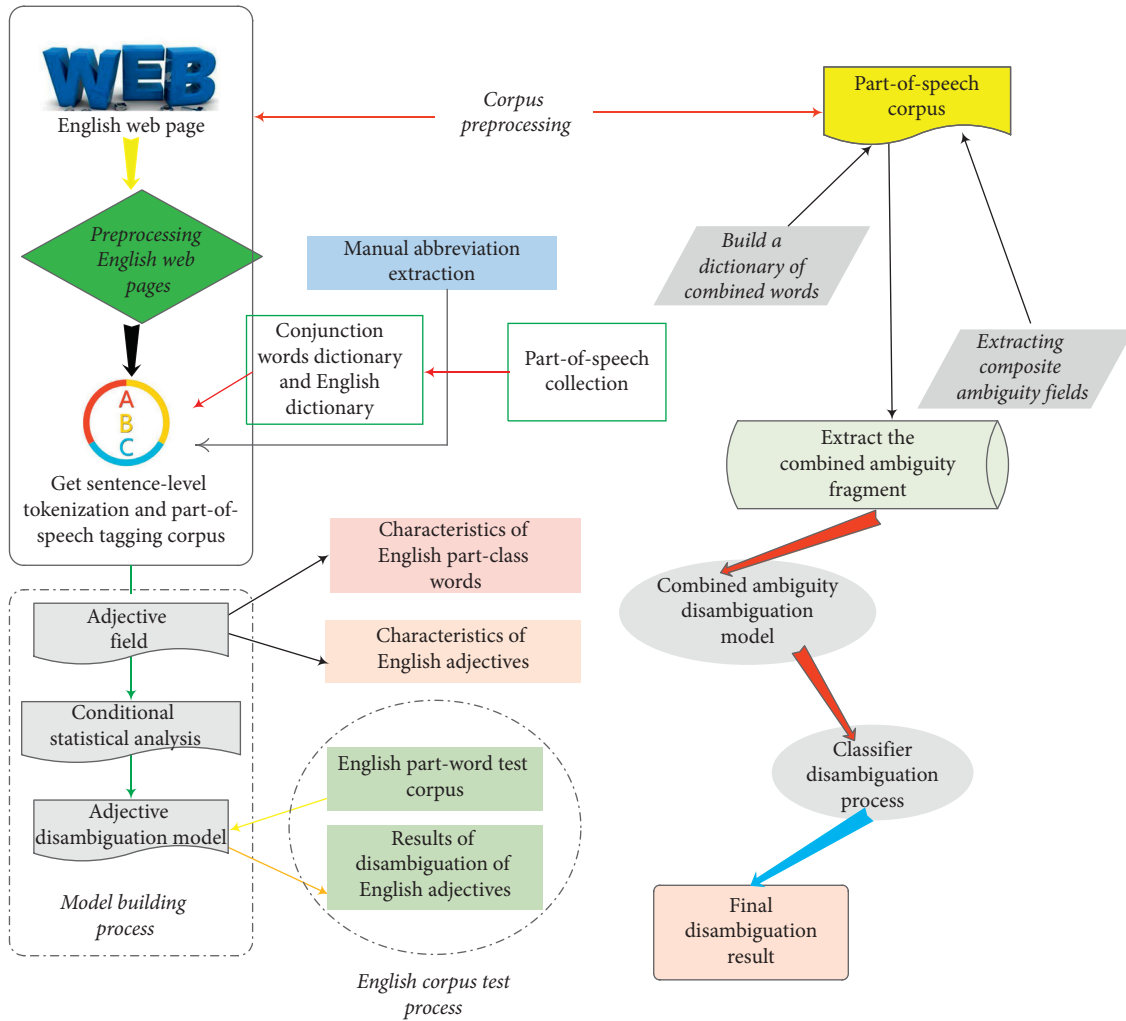


FIGURE 2: English combination ambiguity resolution framework.

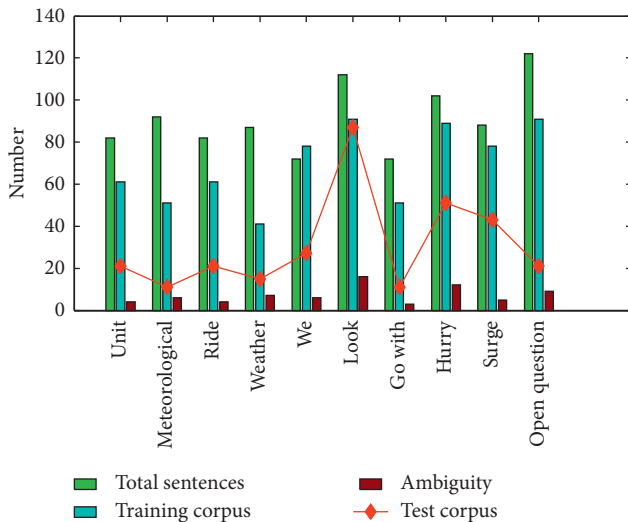


FIGURE 3: Detailed corpus statistics.

The word meaning encoding extraction method obtains four semantic class codes. The first-level encoding extraction method obtains three first-level semantic codes, which are “D, H, and J.” Although “Synonym Cilin” has a more precise five-layer structure, which is the “Synonym Cilin” expansion board, this version of the semantic dictionary is different from the purpose of “Synonym Cilin” at first, but it is closer to the definition dictionary. If the extended version is used for implementation, the three-layer structure code of the word is almost the same as the word form information of the word, which results in no essential difference between the two feature vector sets.

From the above coding set, the most probable semantic knowledge information in the target word window is calculated and obtained according to the Dice coefficient method. Using three different feature extraction methods, three different semantic knowledge information feature vector sets are formed, and a semantic-based disambiguation model is constructed. Finally, the test corpus is used to perform experiments in the disambiguation model, and the accuracy of the models formed by three different feature extraction methods is recorded for the final comparison.

In the experimental stage, in order to measure the impact of training corpus and test corpus distribution on word sense disambiguation, relevant statistics were performed on the training data set. In a given corpus, the number of occurrences of each corpus is counted, and the results are shown in Figure 4.

It can be seen from Figure 4 that the interpretation of each vocabulary of the training corpus is about three times that of the test corpus. This corpus division is also in line with the simple cross-validation in statistical learning. In order to compare the results of the word sense disambiguation method proposed in this paper, we test the efficiency of the method through comparative experiments. Figure 5 lists the disambiguation accuracy of the two different methods.

### 3. Unsupervised Professional Field Segmentation Ambiguity Resolution Method

*3.1. Analysis of Word Segmentation Ambiguity in Professional Field.* The words that cause the ambiguity of the professional domain corpus and the general domain corpus are mainly short words. In the Fish corpus, two-word words accounted for 79%, and three-word words accounted for 13%; in the Bird corpus, two-word words accounted for 81%, and three-word words stood for 14%. Compared with the general field segmentation ambiguity, the composition of professional field segmentation ambiguity is slightly more complicated. In the professional domain corpus, not only ordinary words and ordinary words but also ordinary words and domain words cause segmentation ambiguity. For example, the term “neutralization” (neutralization reaction, the exchange of acids and bases with each other, and the reaction of salt and water) is often used in chemistry. In common sentences, “remaining in the measuring cup and the test strip” is often used.

The professional field lacks labeled training corpus and professional field knowledge, cannot count  $N$ -gram grammatical information, and cannot quickly generate professional field word segmentation ambiguity resolution rules; the professional field words and common words in different sentences cannot be generalized. Based on the above reasons, the general Chinese word segmentation ambiguity resolution method is not suitable for solving the problem of word segmentation ambiguity in professional fields.

Chinese word segmentation mainly includes two types of word segmentation ambiguity: cover ambiguity and overlap ambiguity. Statistical analysis of the number and phenomenon of the two types of ambiguities in the corpus shows that the number of coverage ambiguities is small and the number of overlap ambiguities is large.

Based on the vocabulary, the method of “FMM + back word” can be used to find the overlapped ambiguities that may exist in the test corpus. A set of overlapping strings owned by an overlapping type cut is called an overlapping string chain, and its number is called a chain length. For example, “Secondary” consists of two words overlapping “Second” and “Secondary.” The set of overlapping strings is {“time”}, and the chain length of the overlapping field is 1. In

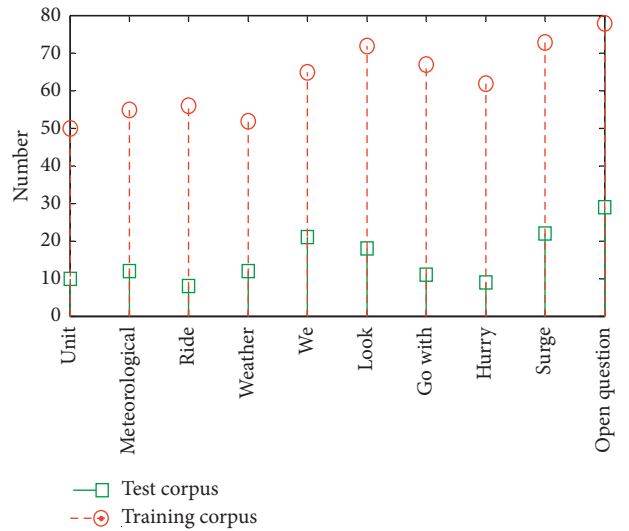


FIGURE 4: Corpus distribution.

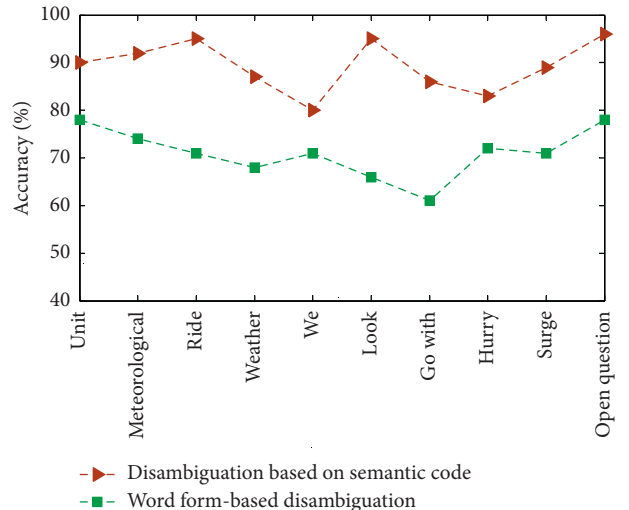


FIGURE 5: Comparison of word sense disambiguation results.

TABLE 1: MOAS chain length statistics.

|              | Fish corpus |      |      |      |      | Bird corpus |      |      |      |      |
|--------------|-------------|------|------|------|------|-------------|------|------|------|------|
| Chain length | 1           | 2    | 3    | 4    | 5    | 1           | 2    | 3    | 4    | 5    |
| Number       | 741         | 269  | 8    | 3    | 2    | 1189        | 422  | 40   | 4    | 2    |
| Share        | 0.73        | 0.26 | 0.01 | 0.02 | 0.01 | 0.72        | 0.25 | 0.02 | 0.02 | 0.01 |

the “molecule,” “combination,” “synthesis,” “component,” and “molecule” all form words, and the chain length of the overlapping field is 3.

Using the word fallback method, based on the Fish vocabulary and Bird vocabulary, we find possible overlap ambiguities in all test corpora and obtain MOAS (Maximal Overlapping Ambiguity String). We classify MOAS in Fish and Bird corpus according to chain length. The statistical results of MOAS chain length are shown in Table 1.



3.2. *Unsupervised Professional Field Segmentation Ambiguity Resolution Method.* Some methods in unsupervised word segmentation are often used to determine the likelihood of a string forming a word in the current corpus. Mutual information can be used to quantitatively estimate the binding force between two Chinese characters. The larger the mutual information, the higher the closeness of the combination of the two Chinese characters; the smaller the mutual information, the lower the tightness of the combination. The calculation formula is

$$\text{mi}(x, y) = \log_2 p(x, y) - \log_2 p(x) - \log_2 p(y). \quad (8)$$

Mutual information can only express the combining power of two Chinese characters. When it encounters a word composed of three or more multicharacters, it cannot express it. Based on mutual information, this paper changes the multiword mutual information. When a word is composed of multiple words, the mutual information of two adjacent words is calculated first, and then the average value is taken. The extended mutual information is

$$\text{mi}'(x_{1,\dots,n}) = (n-1)^{-1} \sum_{i=1}^n \text{mi}(x_i, x_{i+1}). \quad (9)$$

Boundary entropy is also one of the important methods to measure whether a string is a word often used in unsupervised word segmentation. The formula is

$$h(x_{i,\dots,j}) = - \sum_{x \in V} \log p(x | x_{i,\dots,j}) p(x | x_{i,\dots,j}), \quad (10)$$

when  $x$  is  $x_{i-1}$ ,  $x$  is the set of left-adjacent words, and  $h(x_{i,\dots,j})$  is called the left-handed conditional entropy and when  $x$  is  $x_{j+1}$ ,  $x$  is the set of right-adjacent words, and  $h(x_{i,\dots,j})$  is the right-handed conditional entropy. The greater the left and right entropy of a string, the more likely it is a word.

String boundary entropy can disambiguate segmentation in two ways. The first is to judge the possibility of string formation. The second is to judge the degree of separation between words. When judging the possibility of forming a string of words, the left entropy and right entropy of each string in the test corpus are statistically calculated, and the average value of the left entropy and right entropy of each word is taken as a measure. When the string  $abc$  has two splitting methods,  $ab/c$  and  $a/bc$ , the string boundary entropy used by the strings  $ab$  and  $bc$  is

$$\begin{aligned} h_{\text{avg}}(ab) &= 0.5[hr(ab) + hl(ab)], \\ h_{\text{avg}}(bc) &= 0.5[hr(bc) + hl(bc)]. \end{aligned} \quad (11)$$

To determine the degree of separation between strings, we use the following formula:

$$h_{\text{separate}}(x_{i,\dots,j}y_{i,\dots,j}) = 0.5[hr(x_{i,\dots,j}) + hl(y_{i,\dots,j})]. \quad (12)$$

When the string  $abc$  has two splitting methods,  $ab/c$  and  $a/bc$ , the degree of separation between  $ab$  and  $c$  can be expressed as

$$h_{\text{separate}}\left(\frac{ab}{c}\right) = hl(c) + hr(ab). \quad (13)$$

The degree of separation between  $a$  and  $bc$  can be expressed as

$$h_{\text{separate}}\left(\frac{a}{bc}\right) = hl(bc) + hr(a). \quad (14)$$

Boundary entropy can be used not only to judge the possibility of forming a string of multiple words but also to judge the possibility of a single word becoming a “single word.”

In the general field segmentation, the more frequently a word in the training corpus becomes a single word, the more likely it is that the word will become a single word in the test corpus. In the string  $abc$  with a length of 3, whether it is split into  $ab/c$  or  $a/bc$ , you can determine who has  $a$  and a word more frequently in the training corpus; if  $a$  is more frequent than a single word  $c$ , the segmentation result is  $ab/c$ . There is no training corpus for word segmentation in professional fields, and it is impossible to count the probability of single words. When encountering a string  $abc$  with a length of 3, you can use  $\text{havg}$  to determine which word in the test corpus is more likely to become a single word.

### 3.3. Numerical Simulation Experiment

3.3.1. *Experimental Corpus and Settings.* The test corpus in the experiment is the corpus of fish and birds in the Bio-volume of China Encyclopedia. The participle answer of the corpus is marked with Peking University participle standard.

The choice of experimental corpus follows two basic principles: First, it only observes the ambiguity of the segmentation in the domain and eliminates the interference of the OOV (Out of Vocabulary) problem on the effect of the segmentation. All words needed for the test corpus are added to the vocabulary. The second is the maximum possible simulation of actual application. For this experiment, large-scale vocabularies, common domain vocabularies, and network vocabularies are added to the experimental vocabulary.

Taking string frequency judgment as an example, the experiment considers two cases:

- (1) MOAS consists of three characters  $abc$ , which can be divided into  $ab/c$  and  $a/bc$ . If  $\text{Freq}(ab) > \text{Freq}(bc)$ , it is divided into  $ab/c$ .
- (2) MOAS is more than 3 characters, such as  $abcd$ ; if there are two forms of  $abcd$  and  $ab/cd$ , when  $\text{Freq}(abc) > \text{Freq}(cd)$ , it is divided into  $abcd$ ; when  $\text{Freq}(cd) > \text{Freq}(abc)$ , you need to consider again. After dividing  $cd$  into words,  $ab$  is still a word in the dictionary. If  $ab$  does not form a word,  $abcd$  is still used; if  $ab$  is formed, it is  $ab/cd$ .

When mutual information is used as the evaluation criterion, the experimental procedure is the same as the frequency of the character string, and  $\text{Freq}$  is changed to  $\text{mi}$ . When taking the possibility of forming a string boundary entropy as the evaluation criterion, the experimental steps are the same as the string frequency experiments, and  $\text{Freq}$  is changed to  $\text{havg}$ .

TABLE 2: Treatment of 10 experiments.

| Experiment number | Processing situation  |
|-------------------|---|
| 1                 | Baseline segmentation results without any ambiguity   |
| 2                 | Baseline segmentation results without any ambiguity   |
| 3                 | Freqword first uses the vocabulary to segment the test corpus, then counts the word frequency information in the test corpus after segmentation, and finally uses the new statistical word frequency information to segment the test corpus |
| 4                 | Freq only uses string frequency to process ambiguous segmentation results   |
| 5                 | Freq + zi_havg uses a single-word boundary entropy judgment when MOAS consists of 3 characters, and more than 3 characters use Freq judgment  |
| 6                 | mi uses only mutual information   |
| 7                 | mi_zi + havg uses single-word boundary entropy when MOAS is composed of 3 characters, and mutual information is used for more than 3 characters   |
| 8                 | word_havg segmentation results use string boundary entropy to determine the likelihood of word formation  |
| 9                 | word_hseparate uses only string boundary entropy to judge string separation results   |
| 10                | word_havg + zi_havg uses the boundary entropy of a single word when MOAS is composed of 3 characters and uses the boundary entropy of a string to determine more than 3 characters  |

When the separation degree of the string boundary entropy is used as the evaluation standard, the experimental steps are the same as the string frequency experiments, but the evaluation rule is changed to determine hseparate ( $ab/c$ ).

**3.3.2. Experimental Results and Analysis.** In order to analyze only the effect of word segmentation ambiguity, based on the Fish vocabulary and Bird vocabulary without OOV, the FMM word segmentation method was used to test the corpus segmentation of Fish and Bird, respectively. The segmentation result without any ambiguity is used as the baseline of the experiment. The processing conditions corresponding to the 10 experiments are shown in Table 2.

The data in Figures 6 and 7 show that, after only using string frequency to resolve the ambiguity problem, the F1 value of the word segmentation result has been significantly improved, the Fish result has increased by 0.48 percentage points, and the Bird result has increased by nearly 0.9 percentage points. Mutual information and string boundary entropy behave differently on different corpora. Mutual information has the best effect on the Fish corpus, with an increase of nearly 0.8 percentage points; on Bird corpus, the best effect is obtained by mixing the string boundary entropy with the single-word boundary entropy, increasing by 1.2 percentage points. The Fish-based corpus of  $m_i$ -based segmentation surpasses Topline.

The experimental results of word\_hseparate show that using this method to resolve the ambiguity of the word segmentation, the effect of the word segmentation decreases. The reason is that this method is equivalent to a local optimal solution, which results in many erroneous results.

In Bird’s corpus, mixed word boundary entropy is better than single-word entropy. However, the results of the Fish corpus are different. The statistical data of single-word boundary entropy comes from the test corpus. If some words in the test corpus often appear at various positions of the word, the single-word boundary entropy of the word will also be high. For example, the “fish” in the Fish corpus appears very frequently, and the single-word boundary entropy of “fish” is very high, but the possibility of becoming

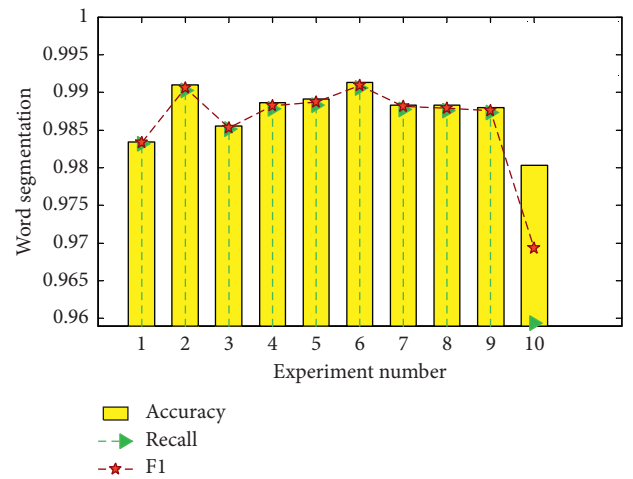


FIGURE 6: Various word segmentation results of Fish without OOV.

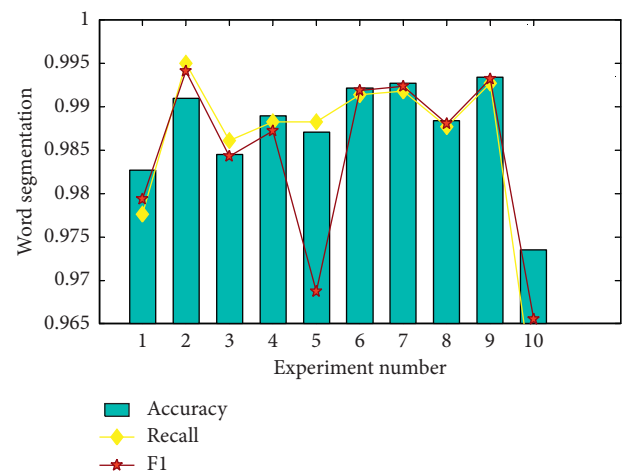


FIGURE 7: Bird’s various word segmentation results without OOV.

a single word is very small. It is generally used as the beginning and end of a word, such as “fish bait” and “fish.” There are many such words in the Fish corpus, and the

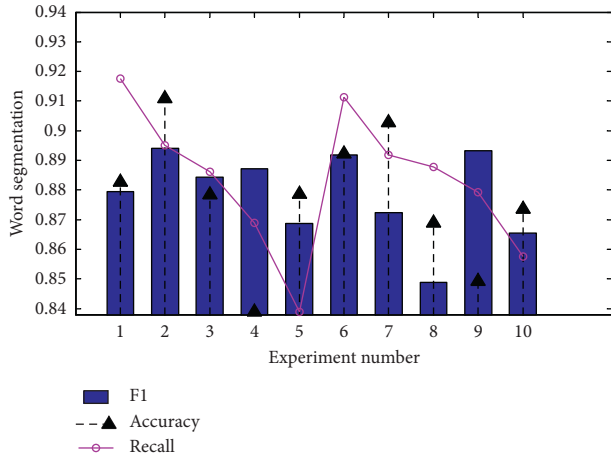


FIGURE 8: Various word segmentation results of Fish with OOV.

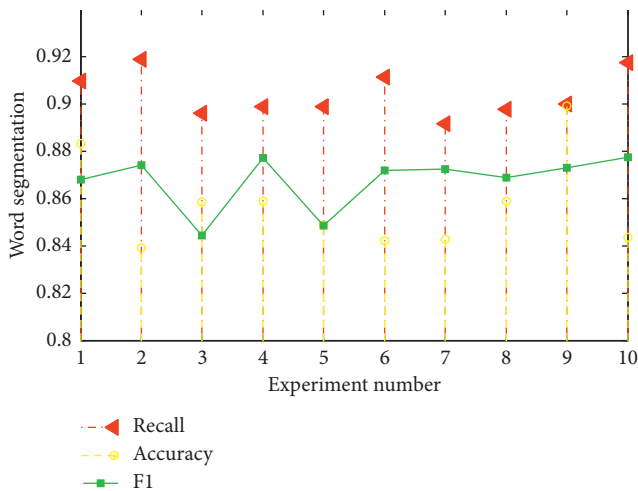


FIGURE 9: Various word segmentation results of Bird with OOV.

single-word boundary entropy does not improve the Fish corpus. There are fewer such words in birds, and the accuracy of the single-word boundary entropy is higher, which has an effect on the result of word segmentation.

The boundary entropy of the word in the test corpus is subtracted from the boundary entropy of the word in the vocabulary, and the difference obtained by subtraction is the final single-element boundary entropy. The boundary entropy of some words after subtraction is negative. In this experiment, the single-word boundary entropy is only used for comparison, it is not used for other operations, and no other processing is performed on negative boundary entropy.

To further verify the validity of the results, the word segmentation vocabulary was changed. The professional words required by the test corpus are halved randomly, so that 879 OOVs (word types) in Fish cannot be recognized, and 1468 OOVs (word types) in Bird cannot be recognized.

The data in Figures 8 and 9 show that, in the presence of OOV, using string frequency, mutual information, and boundary entropy to process ambiguity, the segmentation results are better than baseline. Fish's best results and Bird's

best results are improved. Among all ambiguity processing methods, the performance of mutual information is the most stable.

## 4. Conclusion

According to the constructed feature vector set and through comparison, this paper chooses a naive Bayesian model with better efficiency and accuracy. A word sense disambiguation classifier is established based on the characteristics of semantic codes. A comparison experiment is performed through three different feature vector sets to analyze the performance of the word sense disambiguation classifier based on semantic information. The experimental results show that the accuracy of disambiguation of Bayesian word sense disambiguation model based on semantics is greatly improved compared with the traditional word sense disambiguation model. The general domain segmentation ambiguity resolution method requires professional knowledge and training corpus and is not suitable for solving the professional domain segmentation ambiguity problem. In this paper, the frequency, mutual information, string boundary entropy, and single-word boundary entropy of the test corpus are used as evaluation criteria to solve the problem of word segmentation ambiguity in professional fields. Experiments show that the three evaluation standards can solve the problem of word segmentation ambiguity in professional fields to varying degrees. Among them, the segmentation words using mutual information have the best results and perform stably. The word segmentation method combined with unsupervised learning in the professional field is simple and easy, and it can effectively reduce the problem of word segmentation ambiguity in the test corpus and improve the word segmentation effect in the professional field. However, natural language processing is increasingly a large-scale corpus. Therefore, how to deal with various types of ambiguity fields and improve the speed of word segmentation needs further study.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] L. Deng and D. Li, "Multimedia data stream information mining algorithm based on jointed neural network and soft clustering," *Multimedia Tools and Applications*, vol. 78, no. 4, pp. 4021–4044, 2019.
- [2] W. S. Grant and L. Itti, "Learning invariant features in modulatory networks through conflict and ambiguity," *Neural Computation*, vol. 31, no. 2, pp. 344–387, 2019.
- [3] Y. Graham, T. Baldwin, A. Moffat, and J. Zobel, "Can machine translation systems be evaluated by the crowd alone," *Natural Language Engineering*, vol. 23, no. 1, pp. 3–30, 2017.
- [4] P. Iswarya and V. Radha, "Adapting hybrid machine translation techniques for cross-language text retrieval system,"

- Journal of Engineering Science and Technology*, vol. 12, no. 3, pp. 648–666, 2017.
- [5] M. M. A. Shquier and K. M. Alhawiti, “Fully automated Arabic to English machine translation system: transfer-based approach of AE-TBMT,” *International Journal of Information and Communication Technology*, vol. 10, no. 4, pp. 376–391, 2017.
- [6] M. Howlett, “Moving policy implementation theory forward: a multiple streams/critical juncture approach,” *Public Policy and Administration*, vol. 34, no. 4, pp. 405–430, 2019.
- [7] S. Zannettou, M. Sirivianos, J. Blackburn, and N. Kourtellis, “The web of false information,” *Journal of Data and Information Quality*, vol. 11, no. 3, pp. 1–37, 2019.
- [8] E. M. Garcia, C. Creus, C. España-Bonet, and L. Màrquez, “Using word embeddings to enforce document-level lexical consistency in machine translation,” *The Prague Bulletin of Mathematical Linguistics*, vol. 108, no. 1, pp. 85–96, 2017.
- [9] M. Bouchakwa, Y. Ayadi, and I. Amous, “Multi-level diversification approach of semantic-based image retrieval results,” *Progress in Artificial Intelligence*, vol. 9, no. 1, pp. 1–30, 2020.
- [10] J. Joseph and V. Gaba, “Organizational structure, information processing, and decision-making: a retrospective and road map for research,” *Academy of Management Annals*, vol. 14, no. 1, pp. 267–302, 2020.
- [11] S. Mall and U. C. Jaiswal, “Survey: machine translation for Indian language,” *International Journal of Applied Engineering Research*, vol. 13, no. 1, pp. 202–209, 2018.
- [12] A. Rezapour, S. M. Fakhrahmad, and M. H. Sadreddini, “Applying various distance functions and feature extraction schemes to ambiguity resolution,” *Intelligent Data Analysis*, vol. 22, no. 3, pp. 617–638, 2018.
- [13] Y. F. Hassan, “Rough set machine translation using deep structure and transfer learning,” *Journal of Intelligent & Fuzzy Systems*, vol. 34, no. 6, pp. 4149–4159, 2018.
- [14] P. Xiu and L. Xeauiyin, “Human translation vs machine translation: the practitioner phenomenology,” *Linguistics and Culture Review*, vol. 2, no. 1, pp. 13–23, 2018.
- [15] G. Kotzé, V. Vandeghinste, S. Martens, and J. Tiedemann, “Large aligned treebanks for syntax-based machine translation,” *Language Resources and Evaluation*, vol. 51, no. 2, pp. 249–282, 2017.
- [16] F. Maniez, “An appraisal of recent breakthroughs in machine translation: the case of past participle-based compound adjectives in ESP,” *ASp*, vol. 2017, no. 72, pp. 29–48, 2017.
- [17] A. Benabdallah, M. A. Abderrahim, and M. E.-A. Abderrahim, “Extraction of terms and semantic relationships from Arabic texts for automatic construction of an ontology,” *International Journal of Speech Technology*, vol. 20, no. 2, pp. 289–296, 2017.
- [18] J. Ive, A. Max, and F. Yvon, “Reassessing the proper place of man and machine in translation: a pre-translation scenario,” *Machine Translation*, vol. 32, no. 4, pp. 279–308, 2018.
- [19] M. Kathuria, C. K. Nagpal, and N. Duhan, “A fuzzy logic based synonym resolution approach for automated information retrieval,” *International Journal on Semantic Web and Information Systems*, vol. 14, no. 4, pp. 92–109, 2018.
- [20] K. Cui and X. Jing, “Research on prediction model of geo-technical parameters based on BP neural network,” *Neural Computing and Applications*, vol. 31, no. 12, pp. 8205–8215, 2019.
- [21] M. Chen, S. Lu, and Q. Liu, “Uniform regularity for a Keller-Segel-Navier-stokes system,” *Applied Mathematics Letters*, vol. 107, Article ID 106476, 2020.
- [22] J. Kapočiūtė-Dzikiėnė, A. Davidsonas, and A. Vidugirienė, “Character-based machine learning vs. language modeling for diacritics restoration,” *Information Technology and Control*, vol. 46, no. 4, pp. 508–520, 2017.

## Research Article

# Emotional Interactive Simulation System of English Speech Recognition in Virtual Context

**Dan Li** 

*School of Foreign Languages, Luoyang Institute of Science and Technology, Luoyang, Henan 471023, China*

Correspondence should be addressed to Dan Li; 200901200802@lit.edu.cn

Received 25 May 2020; Accepted 27 July 2020; Published 11 August 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Dan Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of virtual scenes, the degree of simulation and functions of virtual reality have been very complete, providing a new platform and perspective for teaching design. Firstly, the hidden Markov chain model is used to perform emotion recognition on English speech signals. English speech emotion recognition and speech semantic recognition are essentially the same. Hidden Markov style has been widely used in English speech semantic recognition. The experiments of feature extraction and pattern recognition of speech samples prove that Hidden Markovian has higher recognition rate and better recognition effect in speech emotion recognition. Secondly, combining the human pronunciation model and the hearing model, by analyzing the impact of the glottis feature on the human ear hearing-model feature, the research application of the English speech recognition emotion interactive simulation system uses the glottis feature to compensate the human ear, hearing feature is proposed by compensated English speech recognition, and emotion interaction simulation system is used in the English speech emotion experiment, which has obtained a high recognition rate and showed excellent performance.

## 1. Introduction

Situation, virtual situation and instructional design are the starting points of this thesis. The teaching research has been concentrated on the discussion of teaching content and teaching methods, and there is almost no teaching research on the perspective and level of time and space and artistic conception. The research on the situation is also focused on the research of the “situation teaching” or “situation learning” theory and its teaching methods, or the pure situation or concept research is very fragmented, unsystematic, and not combined with educational discussions. In order to deepen the situation-based teaching research, especially the English speech recognition emotion interactive simulation teaching design in virtual reality, it is necessary to study and explore the essence of the situation and virtual situation and significance of teaching, which seems very meaningful.

Usually speech recognition technology includes three parts: preprocessing, feature extraction, and pattern matching [1, 2]. In the preprocessing process, the

prefiltering, pre-emphasis, windowing, and framing of the original voice signal are completed. Feature extraction is one of the key issues of speech recognition technology. In this link, the preprocessed speech will be mapped to the feature space, and the mapping function is the feature extraction algorithm [3, 4]. Another key issue of speech recognition technology is pattern matching, which divides the extracted features into two parts, called training features and test features, respectively. Substitute the training feature into the pattern matching algorithm, train one or several models, and then input the test feature into the trained model to find the model with the highest degree of matching; then, the model is considered to be the representative of the test feature category. The best combination of feature extraction and pattern matching will produce the best recognition effect. Speech synthesis systems are also called text-to-speech conversion systems. As the name implies, they are to convert text into speech [5, 6], which corresponds to speech recognition technology. The purpose of speech synthesis technology is simply to let the computer speak human-like language. In a speech synthesis system, the grammatical

ordering of basic phonemes, that is, the definition of annotated sequences, is the key to synthesizing grammatical rules with high intelligibility [7, 8]. At present, most definitions for labeling sequences are based on the relationship between the current phoneme and the current sentence. If you define the relationship between the front and back sentences and the front and back phonemes and the current phoneme in detail and use these relationships to generate synthesized speech in the synthesis stage, then it is inevitable. It will make the grammatical ordering of phonemes more in line with language habits, resulting in synthetic sentences with higher intelligibility and naturalness [9]. The literature [10, 11] makes a comprehensive and detailed analysis of this theory from the background, knowledge, contextuality of cognition and learning, assumptions, and basic characteristics of contextual learning and contextual cognition. The literature [12, 13] reviews the research of situational cognition and learning theory in the West, summarizes the research history, the meaning, and main content of situational cognition and learning, and points out the future development of the theory direction. The literature [14, 15] conducted many experiments and innovations on Chinese speech emotion recognition. Exploratory experiments on speech emotion recognition include the use of various improved neural network methods, support vector machines, and fuzzy entropy recognition methods. The test results can achieve a recognition rate up to about 75%. The authors in [16, 17] have also conducted fruitful research on feature vector decomposition and fusion of speech emotion signals and feature vector dimensionality reduction, proposed improved classification quadratic functions, and improved Mahalanobis distance and other classification algorithms, and recognized in experimental data the good test performance on the rate. Currently, commonly used speech synthesis methods are parameter synthesis method and waveform stitching synthesis method [18, 19]. The parameter synthesis method is to synthesize speech by adjusting acoustic parameters, which can be simply understood as the reverse process of speech recognition. The parameter synthesis method is relatively mature in theory, but because of the complexity of the algorithm and because the loss of information cannot be effectively compensated, the synthesis effect is not ideal and has a strong machine smell. In the waveform splicing and synthesis system, the basic phonemes are selected from a prerecorded speech database. As long as the database is large enough, it is theoretically possible to splice out any sentences needed [20, 21]. Since the synthesized speech is basic the phonemes are all from the natural original pronunciation of the recorder, so the clarity and naturalness of the synthesized sentences are relatively high. However, it requires high storage capacity and operation speed. In recent years, the HMM-based speech synthesis method has attracted widespread attention. The HMM-based synthesis system can solve the defects of the waveform stitching synthesis method. It models the speech parameters, then uses the sound database data for automatic training, and finally forms a corresponding synthesis system [22]. Emotional speech recognition is an important branch of speech recognition,

which has aroused widespread concern in recent years. Similar to speech recognition, emotional speech recognition can also be divided into three parts: preprocessing, feature extraction, and pattern matching [23, 24]. The difference is that the features here refer specifically to those features that can characterize the emotional state, such as the pitch Rhythm characteristics such as frequency, formant, and short-term average energy. Rhythm feature extraction algorithm is simple, but it is not effective enough for some states with less emotional opposition [25, 26]. Studies [27, 28] in recent years have shown that emotional speech features are no longer a simple combination of the above prosody features. Features based on pronunciation mechanisms are increasingly used in emotional speech recognition and have achieved good results.

The in-depth analysis and discussion of the situation and the virtual situation provide a more in-depth theoretical basis for the development and teaching application of constructivism, provide a more practical plan, and open a new perspective and thinking for the teaching design in virtual reality. First, it introduces in detail the establishment, acquisition, and storage methods of the English speech recognition emotion interaction emotion speech sample library, classifies it according to the requirements of subsequent experiments, and then explains how to affect the preprocessed speech emotion signal. Next, we introduce and compare the main English speech recognition emotion interaction speech emotion feature parameters currently used and how to use the experimental method to obtain the feature parameter fields required in this article. Second is an introduction to the process of speech emotion recognition, including the current major voice emotion recognition methods in the world, comparing their respective characteristics, and then choosing the hidden Markov model method as the voice emotion recognition method in this paper to complete the recognition of virtual human emotion recognition of English speech recognition emotion interactive speech emotion signals. Find the Hidden Markov Model that is most suitable for the purpose of this article, conduct a detailed comparison experiment between discrete Hidden Markov and continuous Hidden Markov Models, draw a comparison of the two advantages and disadvantages, and finally choose continuous Hidden Markov. The model serves as a model for further work.

## 2. Virtual Situation and Emotion Theory of English Speech Recognition

*2.1. Design of English Phonetics Teaching Based on Virtual Situation.* Virtual context provides users with a set of experiences, which is an external experience compared with the internal experiences of people's inner activities (such as joy, anger, sorrow, and joy). Because any external experience may cause people's inner activities, virtual situations can also cause people's inner activities. Therefore, the virtual situation can enable the learner to produce the same or similar inner activities as the real situation. Virtual situation is a phenomenon that can be directly observed and experienced, but it does not have the substantial accessibility as the real

world. Functionally speaking, the experience in the human mind is caused by the virtual situation and the real situation is not direct. When the users in the virtual situation conduct research and reflection, the virtual situation is a possible real reality. Since virtual situations are possible real realities, you can use virtual situations to understand real realities. Here, the virtual situation becomes a virtual intermediary.

The practice of English speech recognition emotional interaction teaching in virtual situations is in line with the constructivist theory and includes four basic attributes of “situation,” “collaboration,” “conversation,” and “meaning construction.” Constructivism believes that personal experience can accelerate the process of knowledge construction and realize the consolidation and externalization of knowledge, and the most convenient virtual situation is the most effective. The virtual situation allows the learner to interact with various existing information, experience different time and space in the learning process, and also be in contact with various parts of the virtual realm. The virtual situation is almost the same as the real situation in terms of simulation degree and function, which also meets the requirements of constructivism that the situation is “highly realistic.” And when participating in the virtual situation, through the avatar mode, it can actively interact with other things and characters, participate in various English speech recognition emotional interactive teaching activities, and conduct collaborative learning to obtain knowledge and achieve the purpose of knowledge construction, as shown in Figure 1.

In virtual situation-based teaching, the learner plays the role of a cognitive subject. Therefore, in virtual situations, English speech recognition is often used for emotional interaction. English speech recognition emotional interaction refers to the support and promotion of learners’ effective learning. Its core is to give full play to the initiative and enthusiasm of students in learning and fully reflect the learner’s cognitive subjective role. Under constructivist thinking, many methods of situation-based teaching have been proposed. Learning in the context of virtual teaching can also draw on these methods. These methods highlight two aspects of collaboration and self-exploration. Constructivists are those who collaborate between teachers and students and between students and students, which is very important for the collection and analysis of data, the formulation and verification of hypotheses, the self-feedback of the learning process, the evaluation of learning results, and the construction of meaning. In the teaching based on the virtual teaching situation, discussion and communication are the main forms of collaboration, but the form of discussion and communication is carried out under the virtual situation. In the process of collaborative learning, teachers first ask questions to arouse students’ “thinking and discussion.” During the discussion, teachers further lead the problems deeper to deepen students’ understanding and at the same time guide students to correct errors and supplement one-sided cognition.

## 2.2. Introduction to the Emotional State of Virtual Situations.

According to the basic sentiment theory, the human emotional state can be divided into several basic types. Let the

emotional state space set  $T = \{t_i | i = 1, 2, \dots, N\}$ , where  $N$  represents the number of emotional states. Use random variables to scroll through emotional states. Let  $Q_i$  be the probability of  $X = t_i$  (take the  $i$ th emotional state) and satisfy

$$\sum_{i=1}^N t_i = 1, \quad 0 \leq t_i \leq 1. \quad (1)$$

In this way, the probability space of emotional states can be expressed as follows:

$$\begin{pmatrix} T \\ Q \end{pmatrix} = \begin{pmatrix} t_1, t_2, \dots, t_N \\ q_1, q_2, \dots, q_N \end{pmatrix}. \quad (2)$$

In English speech emotion speech recognition, some features need to be found, and these features can accurately distinguish different emotional states. This issue involves a variety of disciplines such as psychology, biology, and signal processing. Table 1 compares the results of different acoustic characteristic parameters under six emotional states (“happy,” “angry,” “fear,” “sadness,” “amaze,” and “disgust”).

With the deepening of the research on the classification of English speech emotion interaction, some researchers have proposed a continuous English speech emotion interaction model, which describes emotion in a continuous space, that is, the dimension theory of emotion. Dimension theory regards the conversion process between different emotional states as a continuous linear process in an  $N$ -dimensional space, and the distance between different emotional states in the dimensional space represents the similarity and difference between them. Among the dimensional theory of emotion, the most widely accepted dimensional model is the two-dimensional space of activation evaluation:

- (1) The degree of activation or arousal reflects the active degree of body energy in a certain emotional state
- (2) The degree of evaluation or happiness is based on the separation and activation of positive and negative emotions

Based on the theory of activation evaluation space two-dimensional English speech emotion interaction classification model, as shown in Figure 2, each emotion is distributed at the outer end, and the center of the emotion wheel model is the natural origin, which is a comprehensive. The state of emotion tends to all emotions in the direction, and no emotion is reflected. The closeness of emotions in the same direction determines the intensity of each emotion, and the absolute value of the length of emotion shows the intensity of emotion in this aspect. The emergence of the English speech emotion interaction model introduces the vectorization theory into the classification of emotion models. Any emotion can be represented by a unique vector. According to the vectorization theory, naturally, the amplitude value of the emotion vector is expressed. The intensity of emotion and the angle of the emotion vector show the trend of the type of emotional interaction of English speech in this direction.

Introduce the continuous emotion model from the two-dimensional space to the secondary derived emotional

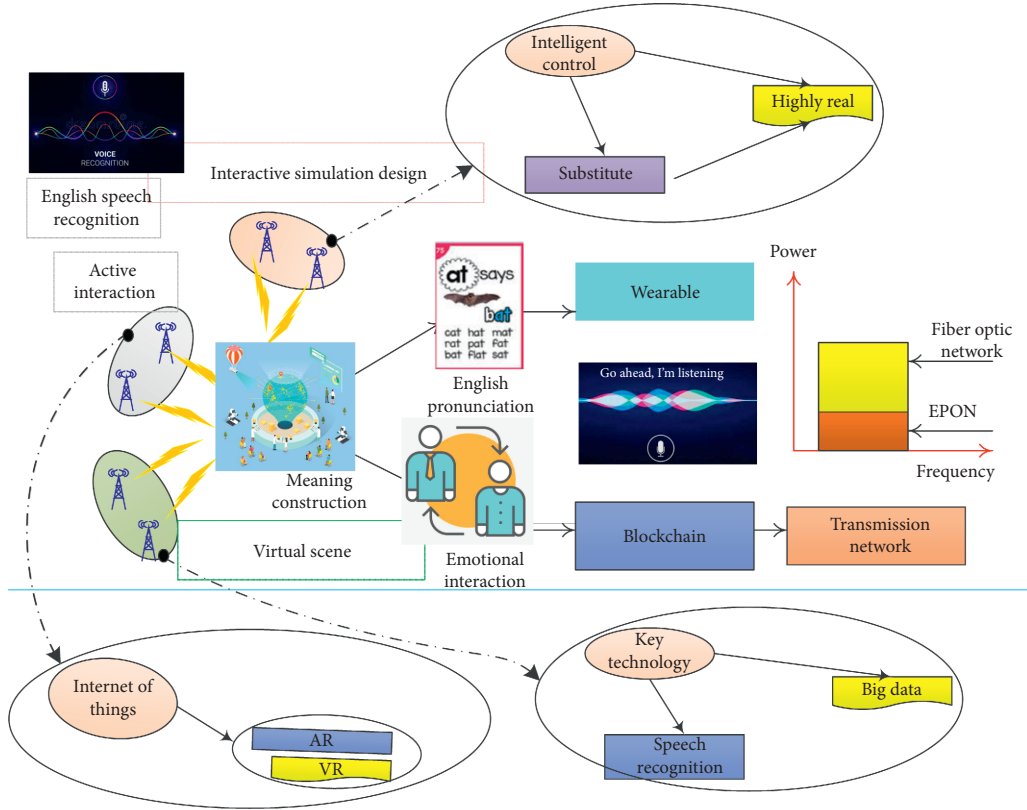


FIGURE 1: English speech recognition emotion interaction mode under virtual scenarios.

TABLE 1: Emotional and speech emotional features.

| Parameter       | Happy     | Pissed off         | Fear              | Sad               | Amazed          | Disgust         |
|-----------------|-----------|--------------------|-------------------|-------------------|-----------------|-----------------|
| Speed of speech | Fast      | Slightly faster    | Quickly           | Slightly slower   | Very fast       | Very fast       |
| Average pitch   | Very high | Very high          | Very high         | Very low          | Very high       | Very low        |
| Pitch range     | Very wide | Very wide          | Slightly narrower | Slightly narrower | Tip up          | Slightly wider  |
| Pitch change    | Smooth    | Mutation at stress | Bend down         | Bend down         | Normal          | Normal          |
| Strength        | Bend up   | High               | Low               | Low               | Irregular sound | Normal          |
| Sound quality   | High      | Breathing          | Resonate          | Resonate          | Accurate        | Irregular sound |
| Sharpness       | Breathing | Chest sound        | Vague             | Vague             | Amazed          | Clear           |

factors, thus introducing the two-dimensional plane space into the three-dimensional space. More importantly, the emotion model in the three-dimensional space combines matrix theory and feature vector. The theory provides the basis of the emotion model for the subsequent vectorized decomposition and fusion of speech emotion feature parameters and the vectorized description of complex emotions in continuous space.

### 3. Research on Emotional Interaction of English Speech Recognition Based on Virtual Situation

**3.1. English Speech Recognition Preprocessing.** In order to convert the continuous analog voice signal into a digital signal for processing and recognition, before extracting the features of the English voice signal, it must go through a series of steps of sampling, digitization, and signal optimization to denoise the voice signal and select recognition

frequency range, high-pass and low-pass filtering, endpoint detection, pre-emphasis, windowing, amplification and control gain, and antialiasing filtering and other voice signal specifications and preparations to facilitate the extraction of voice emotional feature parameters; these tasks are called preprocessing of emotional speech signals.

The English speech signal can be regarded as a non-stationary time-varying signal. Biological research on the principle of sound production by the vocal organs proves that the speed of the state of the vocal organs caused by vibration is much lower than the vibration frequency of the sound. Researchers of emotional signal processing in English speech recognition usually treat emotional signals in English speech recognition as short-term stable signals. In the sound frequency range of 5–50 ms, the spectral characteristics and some physical characteristics of English speech recognition emotion signals are constant. Based on this principle, we have introduced a “windowing” preprocessing method for



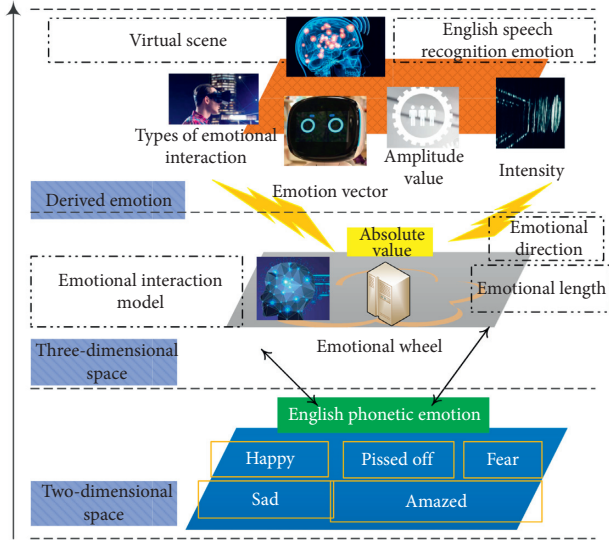


FIGURE 2: English speech emotion interaction classification model.

the English speech recognition emotion signal, combining the short-term processing method of the English speech recognition emotion signal with the processing method and theory of the stable process in a short time slice, and each short-term speech signal segment in a time slice is called an analysis frame. The length of a frame ranges from 10 ms to 30 ms. The “windowing” mentioned by the researchers is a short-term processing method for analyzing frames. The method is the English speech recognition emotion signal in a valid domain which is divided into many short time slices by artificially adding some window functions. The window function of the method of  $w(n)$  to obtain the current frame is to change the sample amplitude outside the divided extraction processing area to zero. The most ideal state of the frequency response of the window function is no side-lobe spectrum leakage, the length is 0, the main-lobe spectrum leakage is close to no, and the length is infinitely narrow. This is only the ideal state that exists in the achieved simulation experiment. The two most commonly used window functions in the rectangular window and Hamming window English speech recognition emotional signal digital processing are

$$w(n) = \begin{cases} 1, & 0 \leq n \leq N-1, \\ 0, & n = \text{else}, \end{cases} \quad (3)$$

$$w(n) = \begin{cases} 0.58 - 0.42 \left( \frac{2\pi n}{N-1} \right), \\ 0, & \text{else}. \end{cases}$$

The characteristic parameters of the speech generated by the linear excitation source filter are mainly reflected in the frequency spectrum structure of speech, in which the formant and pitch frequency are the more commonly used and important characteristic parameters. Formants are one of the most important parameters in speech signal processing.

The difference in the emotional state is reflected in the nervous system to form different degrees of nerve tension, and this difference in tension acts on the vocal muscles, making the same tone. The difference in the degree of muscle tightness is further reflected in the difference in channel frequency. This difference is reflected in the characteristic parameters of the speech signal processing process, which is the formant peak value. A large number of studies have proved that different emotional states will make the position of the formant of the speech signal different. In terms of statistical significance, in order to eliminate the individual differences in formants in the same emotional state on a statistical level, researchers widely use the statistical characteristics of the first three formants peak and bandwidth: average, extreme value, standard deviation, and median, as a research parameter in the process of speech signal processing.

In the extraction of feature parameters of speech signal processing, the linear prediction method based on the all-pole model regards the speech signal  $x(n)$  as a full-pole filter response with  $u(n)$  as the excitation, and from  $u(n)$  to  $x(n)$ , it is expressed as follows:

$$T(z) = \frac{D}{1 - \sum_{i=1}^q b_i z^i} \quad (4)$$

The linear prediction method is used to process the vocal tract model to produce an all-pole model with excellent performance in speech processing research. The practical significance of linear prediction is to decompress the speech signal in the time and frequency domains. The process of deconvolution is to treat the excitation component as the prediction residual  $u(n) = x(n) - \sum_{i=1}^q b_i x(n-1)^i$  and obtain the component of the transfer function  $T(z)$  of the full pole model. The parameter sequence  $[i]$  of this component is further obtained. According to the conclusion that the frequency response characteristic will reach a peak at the resonance peak frequency, the resonance peak is the spectral peak of the frequency response component at this time. By the formula,

$$T(z) = \frac{G}{1 - \sum_{i=1}^q b_i x(n-1)^i} \quad (5)$$

Using the above method, the third formant-related characteristic data in Table 2 below is obtained. From the data in the table, we can think that the third formant has certain distinguishing effects on the four emotions.

For the purpose of feature parameter dimensionality reduction, this paper selects the first 10 numbers from the emotion package of the English speech recognition emotion library for a total of 70 speeches for 7 repetitive trainings, in order to obtain the optimal feature parameter vector combination. In the calculation process, there are two methods for selecting the best feature dimension, one is the cumulative contribution rate method, and the other is to obtain the recognition rate through repeated experiments, and the multiple experiment method of inferring the dimension based on the recognition rate. In the experiment process, only 70 voices were used for dimension reduction

TABLE 2: Statistical parameters related to different emotions of the third formant.

| Third formant frequency | Sad  | Happy | Fear | Pissed off |
|-------------------------|------|-------|------|------------|
| Mean                    | 3048 | 3118  | 3005 | 3067       |
| Maximum                 | 4342 | 4067  | 4098 | 4089       |
| Minimum value           | 2096 | 2106  | 1968 | 2234       |
| Standard deviation      | 356  | 358   | 425  | 354        |
| Median                  | 1908 | 2926  | 2816 | 2945       |

and feature parameter selection, so the method using multiple experiments to find the dimension is more effective. In 7 repeated experiments, the statistical relationship between the average recognition rate and the dimensionality is shown in Figure 3.

As shown in Figure 3, it can be observed that when using feature vector groups with dimensions of 5 to 8, the recognition rate gradually increases; when the dimension exceeds 8, the recognition rate gradually decreases. Using 6 repeated experiments in this paper, the viewpoint of the best vector combination dimension is confirmed so that the conclusion based on the emotional speech library of this article is obtained. The 8-dimensional feature vector group can obtain the best recognition rate. The pitch average, pitch frequency first- and second-order difference, amplitude energy mean, amplitude energy first- and second-order difference, third formant standard deviation, and MFCC standard deviation are eight parameters as speech feature parameters.

### 3.2. English Speech Recognition Based on Virtual Scenarios.

The hidden Markov model is used for emotion recognition. The model characteristics of the double embedded stochastic process characteristics are very consistent with the requirements of emotional state recognition through the emotional feature parameters in speech. The hidden Markov model method is widely used in the field of speech recognition and emotion recognition because of its many statistical advantages as a pattern recognition method; it can be more convenient to train approximate model parameters from a limited number of speech material data. Based on the flexibility of the trained model itself to change materials, optimize the architecture of the cognitive system as the number of training changes and improve the accuracy of the model.

Firstly, a continuous hidden Markov model is used to conduct speech recognition experiments. First, we need to initialize the emotional feature parameters of the speech. The pitch frequency, the first- and second-order difference numbers of the fundamental frequency, the amplitude energy value, the amplitude energy first- and second-order difference values, the first resonance peak, and the eight feature parameters, according to the degree of influence of each feature parameter plus weight value  $w_i$  (where  $\sum_{i=1}^8 w_i = 1$  represents the number of frames) to obtain the eight-dimensional feature vector, are shown below, where  $F_i$  represents the feature vector of each frame and  $i$  represents number of frames:

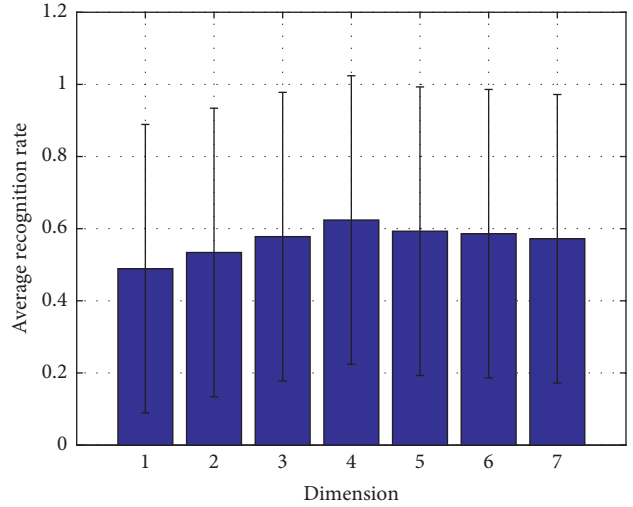


FIGURE 3: Relationship between average recognition rate and dimension.

$$F_i = \left( w_1 \times F_{0i}, w_2 \times \frac{dF_{0i}}{dt}, \dots, w_8 \times F_{1i} \right). \quad (6)$$

The iterative estimation algorithm is used to train the model of speech emotion parameters. The training results of this method are related to the initial value of the parameter (not the initial value of the system). Sometimes, due to the selection of training parameters, two types of situations will occur: the first type, the model cannot converge; in the second category, the converged solution is not the global optimal solution. In order to avoid and solve these two types of situations, we can use the piecewise K-means algorithm to make the training parameters converge better globally optimal solution:

- (1) The input training voice feature parameters are segmented at equal intervals according to the number of states of the HMM model (that is, the number of classifications of the emotion model), and then the feature parameter set in each segment is used as a specific training set to train a certain Emotional state, calculating the initial parameters of the model.
- (2) According to the incoming parameters, the pre-processed voice signal data is divided into the most likely state sequence.
- (3) Use the K-means segmentation algorithm to iteratively re-evaluate the  $B$  in the initial parameters to obtain the intermediate result, that is, collect all the training data of the five emotional states obtained in step (2) to obtain  $B'$ ; if it has been knowing that a model is divided into mixed Gaussian density functions with  $M$  mixed numbers, the K-means clustering algorithm is used to classify all training parameter sets with a state equal to 1 into  $M$  classes and the covariance matrix and mean value of each class. The vector is the center point of the class as the standard point for classification, and finally the mean

estimate  $M$  and variance estimate  $E$  of the  $M$  Gaussian components are regarded as the vector mean and co-square matrix of each type of feature vector set and each Gaussian component. The mixed weight value is as follows:

$$\vartheta_{jm} = \frac{\text{Number of voice frames in state } j \text{ in class } m}{\text{Number of voice frames in } j \text{ state}}. \quad (7)$$

- (4) Substitute the  $\lambda$  obtained in the previous step into the parameter iterative re-estimation method as its initial value re-estimation model, and the re-estimation results in the new model parameter  $\lambda'$ .
- (5) Find the difference between  $\lambda$  and  $\lambda'$ , see if the difference is less than the preset threshold, and see if the template converges; if it does not converge, enter  $\lambda'$  as the new initial parameter and return to step (2). Go to step (6).
- (6) Output the model parameter  $\lambda'$ , that is,  $\lambda'$  is the final model parameter estimation result.

#### 4. Experimental Verification

The selected English speech emotion recognition categories are sad, happy, fear, angry, and neutral. For these five emotional states, 162 groups of speech are selected for each emotion. A total of 810 groups of speech form a training speech database; divided into four times, the model is trained according to the model training volume of 210 groups, 410 groups, 610 groups, and 810 groups. The remaining 190 sets of corpus materials in the 1000 sets are used as the recognition set for the experiment, and the following two sets of tables are obtained.

As can be seen from Table 3, among the 200 groups of unfamiliar emotion speech recognition results, the recognition rate of angry emotion is the highest, reaching 81.2%, the recognition rate of sad emotion is the lowest, reaching 75.4%, and the other three emotions are happy, neutral, and fearful, whose recognition rates are 79.5%, 76.8%, and 77.8%, respectively, and the average recognition rate was 78.122%. Further analysis of the table also shows that the three emotions of sadness, fear, and neutrality are more easily confused in the recognition process. It is speculated that this situation is closely related to the emotional feature parameter domain of these three emotions.

As shown in Table 4, 190 sets of long-term speech and 190 sets of short-term speech were used to perform recognition experiments under 210, 410, 610, and 810 training amounts, respectively. The recognition rate results shown in the table below are obtained.

As shown in Table 4, the following conclusions are drawn. With 200 groups of untrained unfamiliar voices, the recognition rates are 52.53%, 61.96%, 72.49%, and 78.32% under the training volume of 210, 410, 610, and 810 groups, respectively. With the improvement of the training volume, the recognition performance of the resulting model is significantly improved. The recognition ability of the unfamiliar speech shows an upward trend as shown in Figure 4

TABLE 3: Emotion recognition results under 810 training amounts.

| Emotion type | Sad | Happy | Pissed off | Fear | Neutral | Recognition rate |
|--------------|-----|-------|------------|------|---------|------------------|
| Sad          | 115 | 4     | 3          | 20   | 18      | 75.4             |
| Happy        | 5   | 121   | 15         | 13   | 6       | 79.5             |
| Pissed off   | 8   | 12    | 124        | 12   | 4       | 81.2             |
| Fear         | 16  | 11    | 7          | 117  | 9       | 76.8             |
| Neutral      | 14  | 3     | 4          | 20   | 119     | 77.8             |

TABLE 4: Recognition rate of 190 strange recognition corpora in different situations.

| Training volume                        | voice length | 210    | 410    | 610    | 810    |
|--|--------------|--------|--------|--------|--------|
| Long-term voice (3–6 seconds interval) |              | 52.53% | 61.96% | 72.49% | 78.32% |
| Short-term voice (2 seconds)           |              | 53.56% | 65.79% | 84.59% | 89.87% |

below. We can see that although increasing the training volume can recognize the performance, training in 610, the improvement rate of the recognition accuracy rate between the amount of training and the 810 groups of training is not as fast as the stage of 210 groups to 10 groups. Conclusion: increasing the amount of training can improve the accuracy of recognition. The accuracy improvement curve is infinitely close to the extreme value, and there is a bottleneck value in the system recognition rate.

The experimental recognition results of the 210 long-term speech recognition set under the 810 training set are shown in Table 5:

With the increase in training volume, the recognition performance of the resulting model has improved to a certain extent, but the overall recognition performance of the model for long-term speech in this paper is poor; while the model has a significantly higher recognition rate for short-term speech. The recognition rate curve of long-term speech and short-term speech under different training amounts is shown in Figure 5:

As shown in Figure 6, the above five characteristics include not only classic characteristics reflecting the auditory characteristics of the human ear but also recently proposed characteristics reflecting the nonlinear characteristics of the glottis. Through these five characteristics of the recognition experiment of the English emotional speech database and the  $n$  emotional speech database, the comparison of the experimental results' data can directly prove that the English emotional speech database is a more effective emotional speech database, which can be used for future emotional speech research.

As shown in Figure 7, from the perspective of sentiment classification, the performance of LPCC, MFCC, and LPMCC features has changed. In a single speech database experiment, the "neutral" emotion recognition rate is the highest, "happy" is the second, and "angry" is the lowest.

Test in the same training corpus and rule set, observe the changes in the mark recall rate, and mark the accuracy rate. The results are shown in Figure 8.

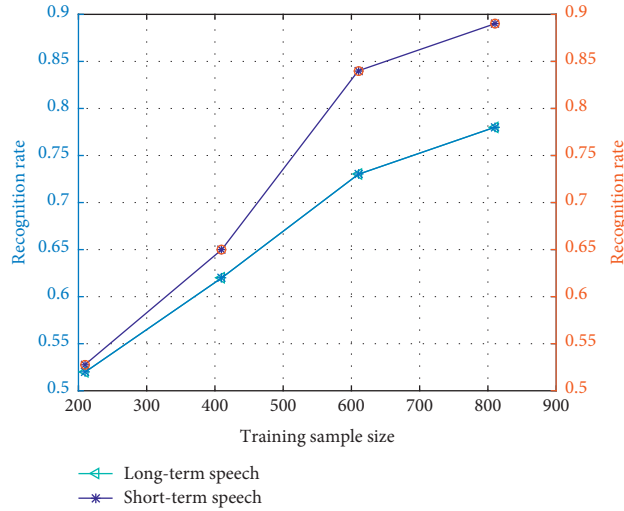


FIGURE 4: Recognition rate of 210 groups of strange voices under different training volumes.

TABLE 5: 210 long-term speech recognition results in the 810 training set.

| Emotion type             | Sad | Happy | Pissed off | Fear | Neutral | Unknown error | Recognition rate (%) |
|--------------------------|-----|-------|------------|------|---------|---------------|----------------------|
| Sad                      | 94  | 9     | 4          | 24   | 23      | 12            | 57.5                 |
| Happy                    | 5   | 103   | 18         | 18   | 6       | 12            | 63.2                 |
| Pissed off               | 11  | 24    | 103        | 18   | 6       | 5             | 63                   |
| Fear                     | 22  | 15    | 14         | 98   | 12      | 8             | 60.5                 |
| Neutral                  | 30  | 11    | 9          | 19   | 88      | 4             | 55.8                 |
| Average recognition rate |     |       |            |      |         |               | 59.995%              |

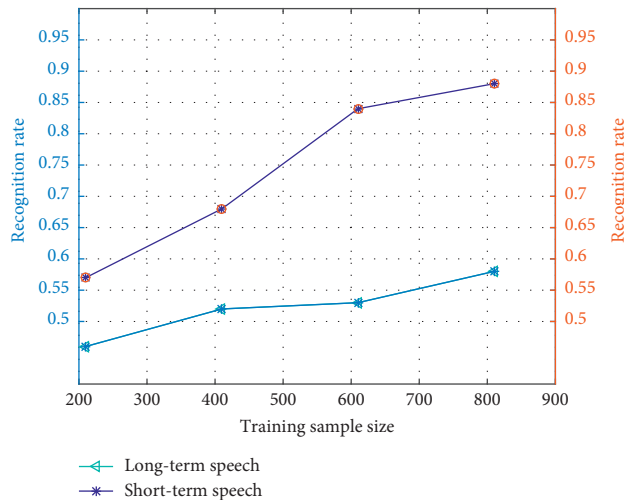


FIGURE 5: Long-term and short-term speech results under different training sets.

It can be seen from Figure 8 that when  $\lambda_1$  takes 0, its LR value is relatively small, and the result is not ideal. Therefore, comprehensive consideration of the rule information and structure co-occurrence information

is indeed very good for improving the accuracy of analysis great help. It can also be seen from the figure that when  $\lambda_2 = 2.5$ , the corresponding LP value is relatively high.

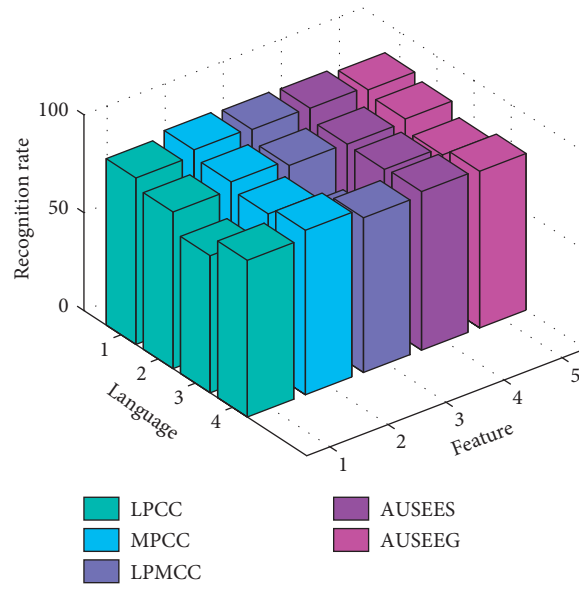


FIGURE 6: Comparison of recognition rates of different features in different languages.

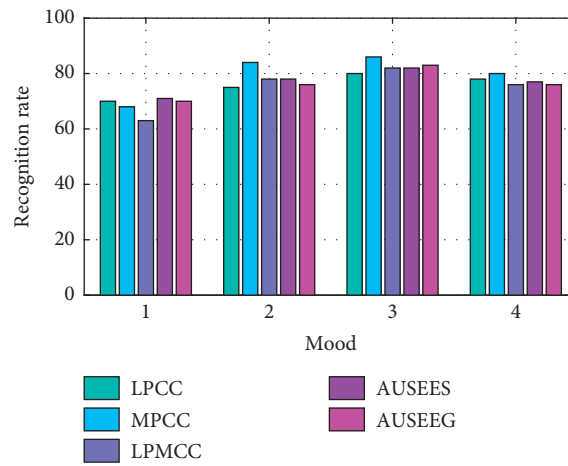


FIGURE 7: Comparison of recognition rates of different features under different emotions.

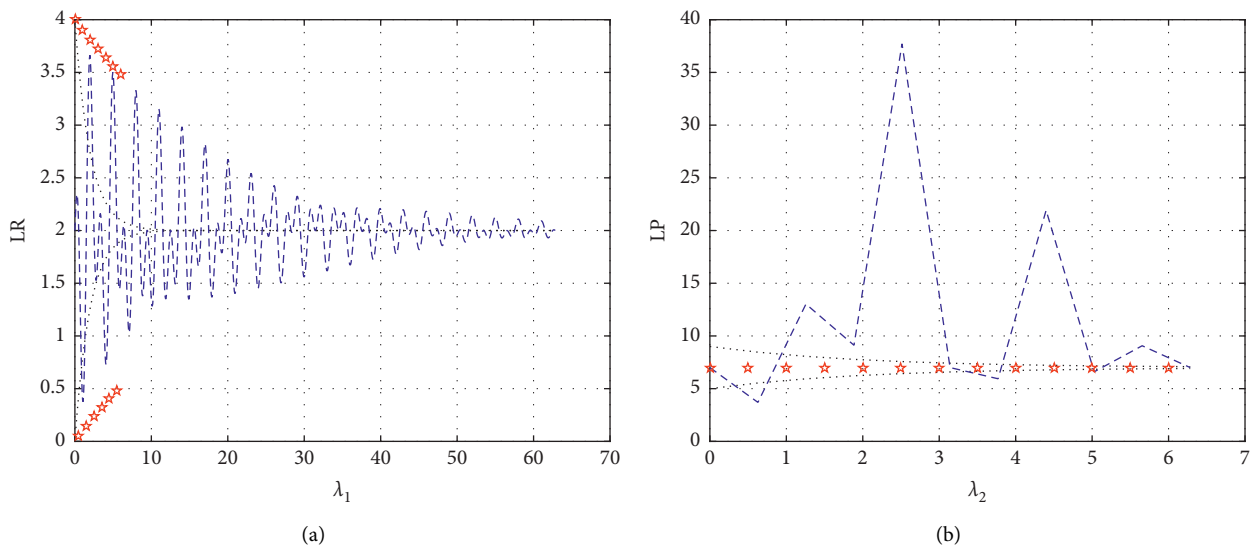


FIGURE 8: Variation of different  $\lambda_1, \lambda_2$  values on the results: (a) the left is LR and (b) the right is the LP change curve.

## 5. Conclusion

This paper synthesizes the research results of pattern recognition and uses 1000 voices of the voice database to train the English voice emotional interaction model in virtual scenarios. Through multiple sets of comparative experiments, the feasibility of the recognition system is verified, and the noise is tested for antijamming capability and robustness. The research work of virtual scene construction adopts the currently popular three-dimensional modeling and three-dimensional virtual reality. Through experiments, the ability of emotional virtual people to recognize and feedback English speech emotion signals in virtual situations is verified. Emotional computing is a highly integrated field of research and technology. By combining computational science, psychology, and cognitive science, we will study the emotional characteristics of human-computer interaction and design a human-computer interaction environment with emotional feedback. It is possible to realize human-machine emotion. With the updating of educational teaching concepts and the advancement of science and technology, there have also been changes in the method of presenting situations. The use of various technical means to present virtual teaching situations has attracted more and more attention from the education community. Therefore, English speech emotional interaction design based on virtual situations will also receive more and more attention.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] Y. Li, X. Lu, C. Yu, H. Guo, and D. Zhang, "Research and application of the virtual simulation system teaching method in NC machining course," *International Journal of Modeling, Simulation and Scientific Computing*, vol. 9, no. 1, Article ID 1850007, 2018.
- [2] N. Hasan, H. Richard, K. Manaf, and B. Fernando, "Basic skin surgery interactive simulation: system description and randomised educational trial," *Advances in Simulation*, vol. 3, no. 1, pp. 14–22, 2018.
- [3] Z. J. Lei, J. J. Huang, Z. Li, L. Wang, J. Cui, and Z. Tang, "Research on collaborative technology in distributed virtual reality system," *Journal of Physics Conference*, vol. 4, no. 2, pp. 960–978, 2018.
- [4] L. Huang, Y.-H. Hou, and D.-J. Zhang, "Research progress on and prospects for virtual brush modeling in digital calligraphy and painting," *Frontiers of Information Technology & Electronic Engineering*, vol. 20, no. 10, pp. 1307–1321, 2019.
- [5] O. Ha and N. Fang, "Effects of interactive computer simulation and animation (CSA) on student learning: a case study involving energy, impulse, and momentum in rigid-body engineering dynamics," *Computer Applications in Engineering Education*, vol. 26, no. 5, pp. 1804–1812, 2018.
- [6] P. K. Ng and B. Tung, "The importance of reward and recognition system in the leadership of virtual project teams: a qualitative research for the financial services sector," *Journal of Transnational Management*, vol. 4, no. 6, pp. 198–214, 2018.
- [7] L. Qin, C. T. Wang, and C. Yao, "Research on application of location technology in 3d virtual environment modelling system for substation switch indicator," *Intelligent Automation and Soft Computing*, vol. 24, no. 1, pp. 115–122, 2018.
- [8] A. R. Sinensis, H. Firman, and M. Muslim, "Reconstruction of collaborative problem solving based learning in thermodynamics with the aid of interactive simulation and derivative games," *Journal of Physics: Conference Series*, vol. 1157, pp. 032042–032054, 2019.
- [9] A.-P. Correia, N. Koehler, and G. Phye, "The application of PhET simulation to teach gas behavior on the submicroscopic level: secondary school students' perceptions," *Research in Science & Technological Education*, vol. 37, no. 2, pp. 193–217, 2019.
- [10] W. V. Bo, G. W. Fulmer, C. K.-E. Lee, and V. D.-T. Chen, "How do secondary science teachers perceive the use of interactive simulations? the affordance in Singapore context," *Journal of Science Education and Technology*, vol. 27, no. 6, pp. 550–565, 2018.
- [11] C. Pratik, A. Akshit, and D. Varun, "Learning in an interactive simulation tool against landslide risks: the role of strength and availability of experiential feedback," *Natural Hazards & Earth System Sciences*, vol. 18, no. 6, pp. 1599–1616, 2018.
- [12] S. Beloufa, J. F. Cauchard, A. KemenyVailleau, F. Mérienne, and J.-M. Boucheix, "Learning eco-driving behaviour in a driving simulator: contribution of instructional videos and interactive guidance system," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 61, no. 2, pp. 201–216, 2019.
- [13] Y. E. Papelis and M. D. Petty, "Recognizing the contributions of reviewers in publishing and peer review," *Simulation*, vol. 94, no. 4, pp. 277–278, 2018.
- [14] H. Cheng, X. D. Wu, and X. Fan, "Modeling and simulation of sheet-metal part deformation in virtual assembly," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 3, pp. 1231–1240, 2019.
- [15] Y. Sang, W. X. Wang, and W. Sun, "Research on the development of an interactive three coordinate measuring machine simulation platform," *Computer Applications in Engineering Education*, vol. 26, no. 5, pp. 1173–1185, 2018.
- [16] X. Xu, Z. Li, L. Wang, and S. Yao, "Interactive visual reality of the offshore hoisting operation and numerical modeling," *International Journal of Pattern Recognition & Artificial Intelligence*, vol. 32, no. 8, p. 1855012, 2018.
- [17] L. Jiangshan and C. Ming, "Research and application of virtual simulation technology in the aerospace bearing design and manufacture," *MATEC Web of Conferences*, vol. 151, no. 8, pp. 04002–04014, 2018.
- [18] L. Yunyue and Y. Fang, "Research and application of paper mill simulation systems based on virtual DPU technology," *Agro Food Industry Hi Tech*, vol. 28, no. 1, pp. 429–432, 2017.
- [19] A. Tilbrook, K. T. Dwyer, and J. A. Parson, "A review of the literature-the use of interactive puppet simulation in nursing education and children's healthcare," *Nurse Education in Practice*, vol. 22, pp. 73–79, 2017.
- [20] A. Kageyama, Y. Tamura, and T. Sato, "Scientific visualization in physics research by CompleXscope CAVE system," *Transactions of the Virtual Reality Society of Japan*, vol. 4, no. 4, pp. 717–722, 2017.

- [21] Y. Peng, Y. Y. Ma, and J. Shan, "The application of interactive dynamic virtual surgical simulation visualization method," *Multimedia Tools and Applications*, vol. 76, no. 23, pp. 25197–25214, 2017.
- [22] B. U. Seeber and S. W. Clapp, "Interactive simulation and free-field auralization of acoustic space with the rtSOFE," *The Journal of the Acoustical Society of America*, vol. 141, no. 5, p. 3974, 2017.
- [23] D. Wang, "Use contexts and usage patterns of interactive case simulation tools by HIV healthcare providers in a statewide online clinical education program," *Studies in Health Technology & Informatics*, vol. 245, pp. 1242–1256, 2017.
- [24] J. Chu, L. Gao, Y. Niu, and G. Li, "Research and application of virtual measuring instrument based on BPNN algorithm in starch industry," *C e Ca*, vol. 42, no. 2, pp. 512–515, 2017.
- [25] B. Qu, Z. Zhang, F. Li et al., "Research and application of oil operator training system based on virtual reality technology," *Journal of Petrochemical Universities*, vol. 30, no. 1, pp. 54–59, 2017.
- [26] M. Y. Kim, Y. Lee, and D. Lee, "Haptic rendering and interactive simulation using passive midpoint integration," *The International Journal of Robotics Research*, vol. 36, no. 12, pp. 1341–1362, 2017.
- [27] Y. S. Lee, H. J. Yap, and R. Singh, "Implementation of a voice-control system for issuing commands in a virtual manufacturing simulation process," *Advanced Materials Research*, vol. 980, pp. 165–171, 2014.
- [28] K. Mcmanus, N. R. Mitchell, and N. Tracy-Ventura, "Understanding insertion and integration in a study abroad context: the case of english-speaking sojourners in France," *Revue française De linguistique Appliquée*, vol. XIX, no. 2, pp. 97–116, 2014.

## Research Article

# Explosion Resistance of Three-Dimensional Mesoscopic Model of Complex Closed-Cell Aluminum Foam Sandwich Structure Based on Random Generation Algorithm

Zhen Wang, Wen Bin Gu , Xing Bo Xie, Qi Yuan, Yu Tian Chen, and Tao Jiang

Army Engineering University of PLA, Nanjing 210007, China

Correspondence should be addressed to Wen Bin Gu; 1120122090@bit.edu.cn

Received 1 June 2020; Accepted 11 July 2020; Published 29 July 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Zhen Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

According to the randomness of the spatial distribution and shape of the internal cells of closed-cell foam aluminum and based on the Voronoi algorithm, we use ABAQUS to model the random polyhedrons of pore cells firstly. Then, the algorithm of generating aluminum foam with random pore size and random wall thickness is written by Python and Fortran, and the mesh model of random polyhedral particles and random wall thickness was established by the algorithm read in by TrueGrid software. Finally, the mesh model is imported into the LS-DYNA software to remove the random polyhedron part of the pore cell. Compared with the results of scanning electron microscopy and antiknock test, the morphology and properties of the model are close to those of the real aluminum foam material, and the coincidence degree is more than 91.4%. By means of numerical simulation, the mechanism of the wall deformation, destruction of closed-cell aluminum foams, and the rapid attenuation of explosion stress wave after the interference of reflection and transmission of bubbles were studied and revealed. It is found that aluminum foam deformation can be divided into four areas: collapse area, fracture area, plastic deformation area, and elastic deformation region. Therefore, the explosion resistance is directly related to the cell wall thickness and bubble size, and there is an optimal porosity rule for aluminum foam antiknock performance.

## 1. Introduction

Closed-cell aluminum foams are a kind of porous metal material, which is composed of thousands of random 3D polyhedral pores embedded in continuous aluminum or aluminum alloy matrix. Compared with other composites, closed-cell aluminum foam has the characteristics of lightweight, high strength, and dual physical properties of function and structure. Because of its unique mesostructural characteristics, the material can have long and almost constant platform stress in compression, which is very beneficial for energy absorption [1, 2]. Therefore, aluminum foams are being increasingly used in energy-absorbing structures [3–11]. It has been widely used in the field of explosion and impact protection. Goel et al. [12] studied the effect of closed-cell aluminum foam on the shock wave passing through the impact tube, and it is found that the existence of aluminum foam interlayer has a great influence

on the reflected wave. According to the research of Shen et al. [13], they carried out the explosion loading experiment on the aluminum foam sandwich plate which was bent outwards and found that compared with the plane aluminum foam sandwich plate, this structure changed the incident angle of detonation wave and the deformation mechanism of the panel. Its antiknock ability was better than the latter. Jing et al. [14–16] introduced experimental studies on antiknock performance of circular arc foam metal composite panels, and the deformation and failure characteristics of such composite plates subjected to blast loading are analyzed.

As we all know, the local failure and deformation of aluminum foam are very serious under the condition of a strong dynamic load. It is difficult to obtain the local mesofailure mode and the failure process (mode) of deformation, collapse, and fracture in the experimental study. Some studies have shown that the mechanical properties of



aluminum foam materials show obvious multiscale characteristics [17]. At the mesolevel, the mechanical behaviors of the aluminum foam cell wall, such as plastic deformation, buckling, and fracture, have a great influence on the macro-mechanical properties. Therefore, scholars at home and abroad have established a large number of numerical models and carried out a variety of research methods [18–23].

At present, there are three main methods to study the mesoscopic model of porous metal materials. The first method adopts repeating unit cell (RUC). The macro-mechanical properties and deformation characteristics of aluminum foam structure are simulated by regularly repeating the predesigned representative unit. The choice of representative units is more diversified. Gibson and Ashby [24] proposed a cubic structure model for closed-cell foam materials. It assumes that the cell structure is a simple and repeated arrangement of cubes, the edge length, edge wall, and wall thickness of each cube are the same, and the elastic modulus and yield strength expression of the foam material are given by this model. In order to more realistically simulate the cell structure of actual aluminum foam, the basic unit shape of this representative model has also been gradually improved. Representative models of closed-cell aluminum foam have successively developed such ideal models as the Kalman model (tetrahedral model) [20, 25, 26], octahedral model [27], more complex cube to pyramid model [28], cube to sphere model [28, 29], all of which use uniform pores, cell walls, and prisms. The main drawback of this method is that it cannot reflect the randomness of the microstructure of aluminum foam. The second type of method considers the randomness of the microstructure of aluminum foam on this basis and generates a cell structure to simulate the pore structure of aluminum foam by certain rules. Many scholars [19, 31, 32] proposed a three-dimensional Voronoi algorithm technology to establish foam three-dimensional mesoscopic model of aluminum. This model can reflect the mesomechanical properties of aluminum foam more realistically, and it has great significance for the study of its energy dissipation mechanism and deformation mode. However, most of the hole walls use shell elements, and their thickness is the same at any position, which is obviously not consistent with the experimental observation results. The third type of method is to perform three-dimensional reconstruction based on the CT scan image of the material to obtain the mesofinite element model [3], which can truly restore the mesoscopic structural characteristics of aluminum foam. However, the number of finite element model elements obtained by this method is huge, and the calculation cost is far more than the previous two methods. In order to more realistically simulate the deformation and failure process of aluminum foam under external load, it is necessary to establish a cell wall model of aluminum foam in line with the actual situation to study the mechanical properties of aluminum foam. Fang et al. [33, 34] used a three-dimensional random polyhedron model algorithm to simulate foam metal materials. The pores were simulated using three-dimensional random convex polyhedrons. At the same time, a random control algorithm for wall thickness was proposed to achieve random distribution

of pores and wall thickness but also increased the calculation of the amount. It is difficult to increase the porosity by a large margin, and there is no real explosive simulation.

In this paper, on the basis of the Voronoi algorithm, the process of random pore size and cell wall thickness is realized by self-programming. A closed-cell aluminum foams modeling method based on a three-dimensional mesomodel is proposed. This method greatly improves the calculation efficiency and shortens the calculation time. At the same time, the sandwich structure of the aluminum foam is designed, and the correctness of the three-dimensional micromodel of aluminum foam is verified by experiments. On the basis of the established finite element model of real structure, the deformation mode, energy absorption effect, and attenuation mechanism of explosion shock wave of aluminum foam sandwich structure were studied, and the wave attenuation law of aluminum foam to explosion loads with different porosities was obtained. These experimental results can provide a theoretical basis for the design of the weight reduction device.

## 2. Mesoscopic Model Setup

The microstructure of closed-cell foam aluminum has a large number of randomly distributed closed cells. In this chapter, we consider the cells in the foam, give the generation algorithm of the closed-cell foam aluminum three-dimensional model, and establish the three-dimensional mesoscopic model of the aluminum foam. The algorithm consists of three steps. The first is the generation of random polyhedron particles. The pores of aluminum foam are regarded as random polyhedral particles. Based on the Voronoi algorithm, ABAQUS is used to model the random polyhedrons and extract their geometric features. The second is to comprehensively use the Python language and Fortran language to write a three-dimensional random aluminum pore size and random wall thickness foam aluminum generation algorithm, so that the aluminum foam cell wall thickness is random, through the TrueGrid software reading algorithm, the establishment of random polyhedron particles, and random wall thickness grid model. The third is to import the mesh model into the LS-DYNA software, random polyhedron particle part of pore cell is removed, generate a foam aluminum model, and then build an explosion model of the foam aluminum sandwich panel.

*2.1. Describe the Randomness of Pore Size and Cell Wall Thickness.* The closed-cell foam metal is composed of a large number of randomly distributed closed cells and cell walls. The pore size is generally 0.5–4 mm, and the cell wall thickness is about 0.05–2 mm. The cell pore size and cell wall thickness are randomly distributed. Metal foams exhibit typical heterogeneity derived from the microstructure of the cell wall.

In this paper, the foam aluminum model algorithm is based on the Voronoi algorithm, using the Python language and Fortran language comprehensively. According to the

mesostructure characteristics of aluminum foam, the modeling method of complex polyhedron is partially improved to Voronoi algorithm. Consider randomness. It not only realizes the randomness of the pore size of the aluminum foam but also realizes the randomness of the wall thickness, so it is a more realistic mesoscopic model of the aluminum foam. The generation of cell walls and particles is shown in Figure 1:

According to the Voronoi algorithm, the discrete data points are reasonably connected to construct a Delaunay triangle network, and then the vertical bisectors of the line segments of two adjacent points are connected to form a Tyson polygon [34], as shown in the figure. Each vertex of the Tyson polygon is the circumscribed circle center of the adjacent triangle. The Tyson polygon is regarded as the pores of two-dimensional aluminum foam. By adjusting the number of polygons, the polygonal pore diameter is controlled within 1–3 mm, and the average pore diameter is 2 mm. Using the Python language and Fortran language comprehensively, a random algorithm of random cell wall thickness is written to obtain the cell wall as shown in Figure 1(b), and the size of the pore size is finely adjusted by the cell wall thickness. At this point, the geometric model of the two-dimensional aluminum foam and the cell wall is generated. After the algorithm is improved, it can be programmed with shorter sentences, which is highly executable, shortens the modeling running time, and generates models quickly.

Then, the two-dimensional model is extended to the three-dimensional cube domain, random polyhedrons are generated in the cubes of arbitrary shape and structure as shown in Figure 2, random thickness is given to any geometric surface by random algorithm, then the random thickness of the cell wall the value range is 0.03~0.4 mm, and the random function of the cell wall is as follows:

$$T_d = T_{\max} - (T_{\max} - T_{\min}) \times \zeta_{\text{random}}, \quad (1)$$

where  $T_d$  is the cell wall thickness,  $T_{\min}$  and  $T_{\max}$  are the minimum and maximum cell wall thickness,  $\zeta_{\text{random}}$  is the random distribution function of the cell wall thickness, and the value range is 0~1. A suitable foam cell mesoscopic model should take into account the randomness of non-uniformity, pore size, and cell wall thickness.  $T_{\min}$  and  $T_{\max}$  were taken as 0.03 and 0.4 mm, respectively, until the cell wall thickness satisfies 0.03~0.4 mm to stop the regeneration of random polyhedron and finally get the three-dimensional polyhedron geometric model of Figure 2(b). Compared with other algorithms, this algorithm can effectively shorten the modeling time and improve the calculation efficiency and ensure the randomness of the cell wall and pore size.

**2.2. Generation of Grid.** Use TrueGrid software to read in the random pores and random cell wall foam aluminum algorithm, extract geometric features, and use the mapping grid method to build a grid model of random polyhedral particles and random wall thickness. Aiming at the characteristics of the pore delivery area of the closed-cell foam aluminum mesomodel, a structured grid is divided, and the

characteristic size of the unit is determined according to the pore diameter and wall thickness. In order to take into account both the computational efficiency and the accuracy of the simulation, this paper takes the feature size of the cell as the grid size is set to 0.2 mm. The regular cell distribution characteristics ensure the calculation efficiency of the subsequent material judgment and are conducive to programming. A spatial eight-node hexahedral unit is used to divide the overall delivery area into a uniform grid to obtain a regular initial grid structure. A fine and uniform grid improves computational efficiency while ensuring accuracy. First build a cube model that surrounds the cylinder, as shown in Figure 3(a); then fill the extracted geometric features in Figure 3(a) to get Figure 3(b); according to the position of the grid in the sample, determine the material properties of the grid; when all nodes of the element are located in a random polyhedron, the material property is set to pore; otherwise, it is defined as aluminum material. Then perform cylindrical cutting to delete the aluminum foam model outside the cylinder, and the remaining part is shown in Figure 3(c); and then propose removing the particle model part and obtaining the mesh model of the foam aluminum as shown in Figure 3(d), the output grid file, the three-dimensional view of aluminum foam is shown in Figure 3(e). To facilitate the observation of the internal structure of aluminum foam, a quarter model is intercepted, as shown in Figure 3(f).

**2.3. Modeling of Aluminum Foam Sandwich Structure.** Use TrueGrid software to establish the air grid model and fill the air grid with explosives and the explosive and the air grid share nodes; use SCDM software to establish the geometric model of the sleeve, cover plate, and bottom plate; divide and export the  $K$  files of all fluid and solid grids; assemble the grids in the LS-PrePost software together with the  $K$  files of the aluminum foam; and finally obtain the model and  $K$  file of the aluminum foam sandwich structure. The quarter model and dimensions are shown in Figure 4. The size of the air domain is 0.82 m × 0.42 m × 0.4 mm, the radius of the explosive is 0.043 m, the thickness of the cover plate is 0.01 m, and the thickness of the bottom plate is 0.02 m.

Add a nonreflective boundary condition on the outer surface of the air layer, the bottom of the bottom plate is set as a fixed boundary, the cover plate and the bottom plate are bound and connected, \*CONTACT\_TIED\_SURFACE\_TO\_SURFACE, the cover plate and the aluminum foam, and the aluminum foam and the bottom plate take automatic surface-to-surface contact \*CONTACT\_AUTOMATIC\_SURFACE\_TO\_SURFACE.

In order to accurately analyze the nonlinear behavior of the foam aluminum sandwich panel under strong dynamic load, the system bottom plate, sleeve, bracket, and end cover are Q235 steel. The above materials and foam aluminum are all used in the PLASTIC KINEMATIC material model of LS-DYNA. The metal materials were numerically simulated. The calculation parameters of metal materials are shown in Table 1.

The air uses the \*MAT\_NULL material model, and the equation of state is described using \*EOS\_LINEAR\_

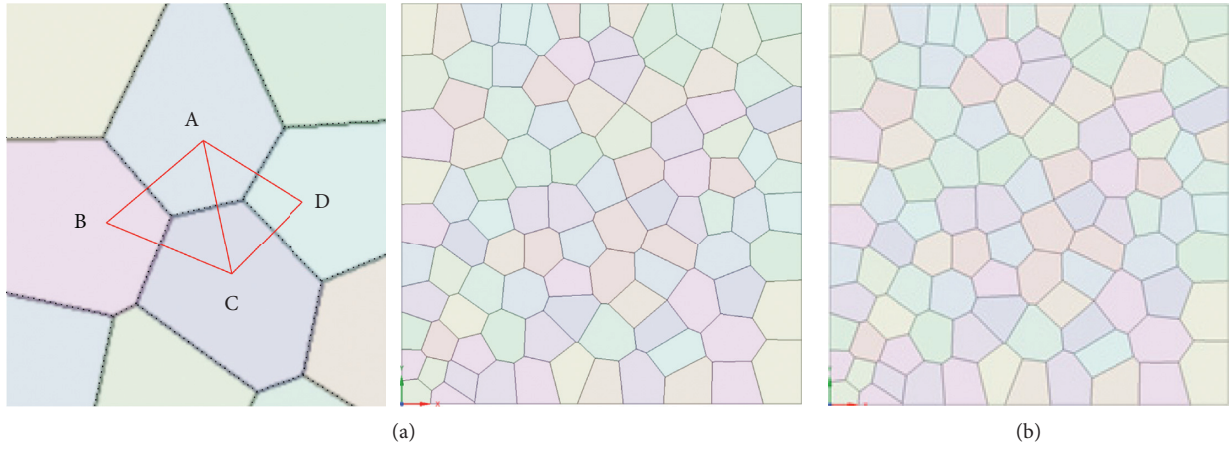


FIGURE 1: Two-dimensional particle and cell wall generation process.

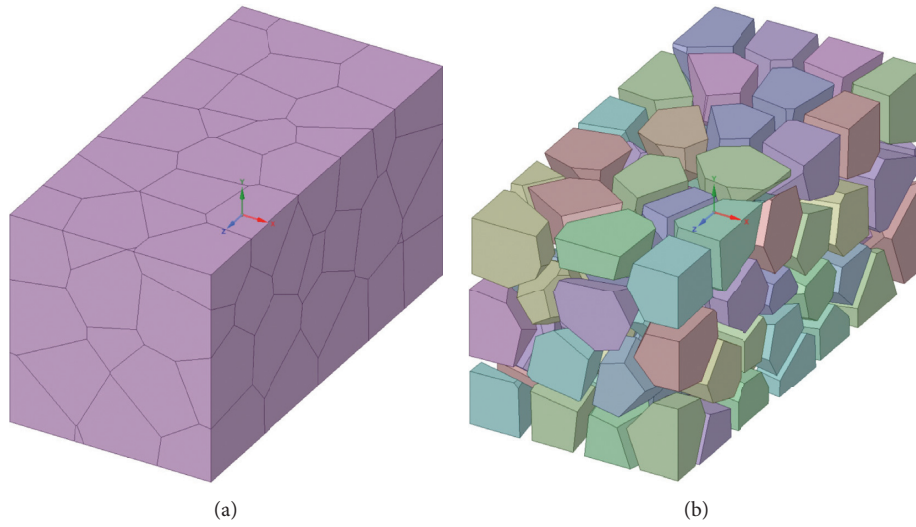


FIGURE 2: Random polyhedron and cell wall generation process. (a) Random polyhedron generation. (b) Generation of three-dimensional cell wall model.

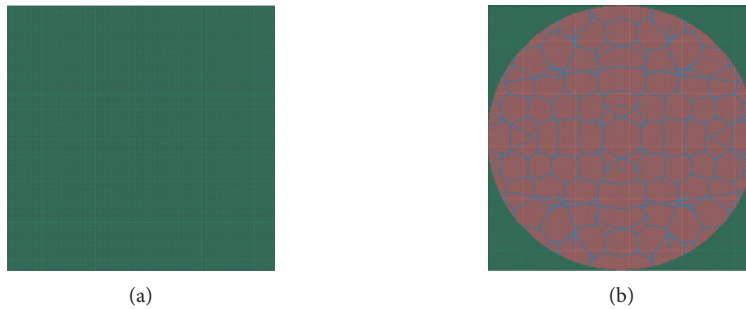


FIGURE 3: Continued.

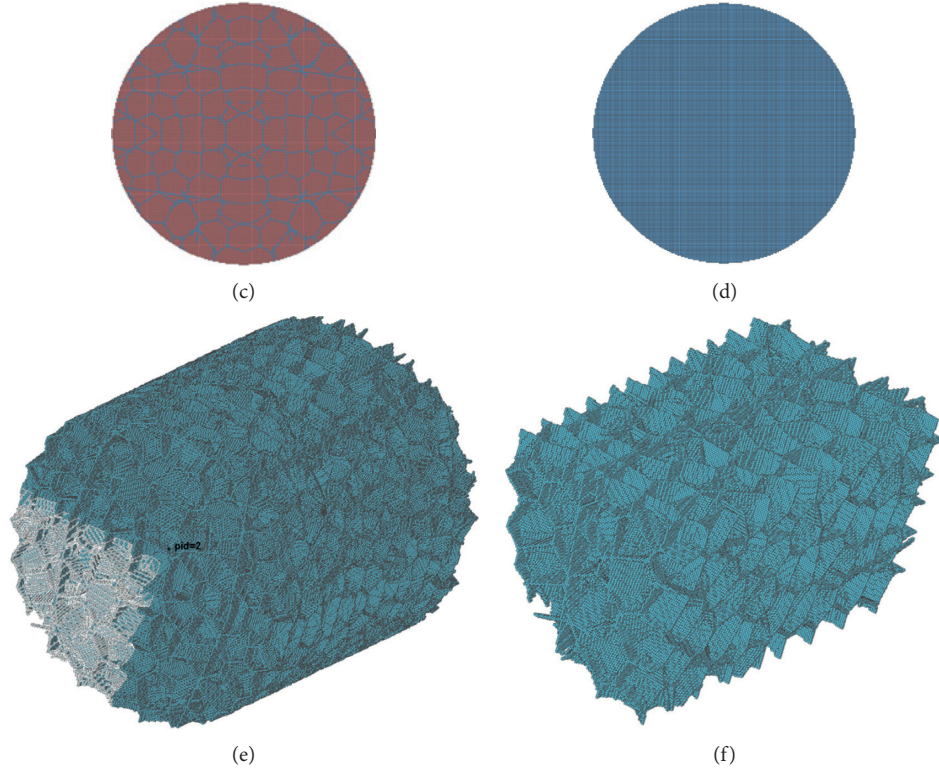


FIGURE 3: Finite element model generation of aluminum foam.

TABLE 1: Metal material parameters.

| Metallic material  | $Q$ (kg/m <sup>3</sup> ) | $E$ (GPa) | $I$   | SIGY (MPa) | ETAN (GPa) |
|--------------------|--------------------------|-----------|-------|------------|------------|
| Q235 steel [35]    | 7830                     | 210       | 0.274 | 235        | 6.1        |
| Aluminum foam [33] | 2730                     | 70        | 0.34  | 185        | 4.62       |

POLYNOMIAL. The expression of the equation of state is as follows:

$$p = C_0 + C_1\mu + C_2\mu^2 + C_3\mu^3 + (C_4 + C_5\mu + C_6\mu^2)E,$$

$$\mu = \frac{1}{V} - 1 = \frac{\rho}{\rho_0} - 1, \quad (2)$$

$$C_4 = C_5 = \gamma - 1.$$

Considering air as a gas in an ideal state, the coefficients of the polynomial equation  $C_0 = C_1 = C_2 = C_3 = C_6 = 0$ . The variable coefficient  $\gamma$  is often set to 1.4, so  $C_4 = C_5 = 0.4$ .  $E_0$ ,  $\rho_0$ , and  $V_0$  which are the initial energy density, initial density, and initial relative volume parameter values are 1.29 g·cm<sup>-3</sup>, 0.25 MPa, and 1.0. The parameters are shown in Table 2.

The high-energy combustion explosives use \*MAT\_HIGH\_EXPLOSIVE\_BURN material model, and the equation of state uses \*EOS\_JWL to represent the pressure of the explosive product. The expression of the equation of state is as follows:

$$p = A \left( 1 - \frac{\omega}{R_1 V} \right) e^{-R_1 V} + B \left( 1 - \frac{\omega}{R_2 V} \right) e^{-R_2 V} + \frac{\omega E}{V}, \quad (3)$$

TABLE 2: Air material parameters [22].

| $C_0$ (MPa) | $C_1$ | $C_2$ | $C_3$ | $C_4$ | $C_5$ | $C_6$ | $\rho$ (kg/m <sup>3</sup> ) |
|-------------|-------|-------|-------|-------|-------|-------|-----------------------------|
| -0.1        | 0     | 0     | 0     | 0.4   | 0.4   | 0     | 1.225                       |

where  $E$  and  $V$  are energy density and relative volume, respectively, and the initial values should be assigned to  $E_0$ ;  $V_0$ ,  $A$ ,  $B$ , and  $E_0$  are pressure units; and  $R_1$ ,  $R_2$ , OMEG, and  $V_0$  are dimensionless. The specific material parameters are shown in Table 3.

And use the ALE algorithm to solve \*CONTROL\_ALE, the explosive and air form an Euler multimatter group (\*ALE\_MULTI-MATERIAL\_GROUP), and the fluid domain and solid domain are set by fluid-solid coupling (\*CONSTRAINED\_LAGRANGE\_IN\_SOLID). Finally submit ANSYS/LS-DYNA for solution.

In the geometric nonlinear analysis of explosion impact, the material often deforms greatly. In order to better fit the changing shape of the aluminum foam cell wall under real conditions, the material erosion method is often used to deal with the distorted unit. When the stress or strain reaches the erosion failure condition, we consider this element to fail and remove it from the model. The failure criterion of

TABLE 3: Explosive material parameters [36].

| $\rho$ (kg/m <sup>3</sup> ) | $A$ (GPa) | $B$ (GPa) | $R_1$ | $R_2$ | PCJ (GPa) | $E$ (J/m <sup>3</sup> ) | $D$ (m/s) | $\nu$ | $\omega$ |
|-----------------------------|-----------|-----------|-------|-------|-----------|-------------------------|-----------|-------|----------|
| 1500                        | 347.6     | 3.39      | 4.15  | 0.95  | 17.92     | 6.34e9                  | 6957.2    | 1.0   | 0.28     |

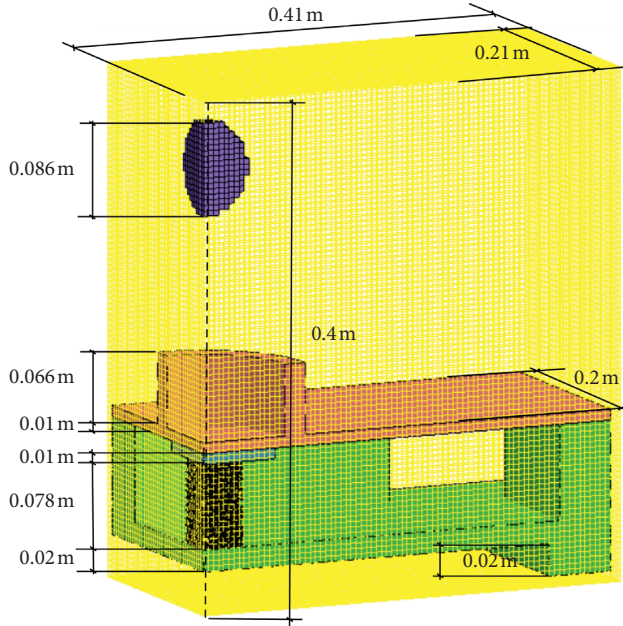


FIGURE 4: Sandwich plate model.

aluminum foam adopts the maximum strain failure criterion. According to the research results, the maximum failure strain of aluminum foam in this paper is 0.37 [37].

### 3. Validation

**3.1. Strain Results Verification.** The antiexplosion performance of the foam aluminum sandwich board was studied. All tests were conducted in the structure in Figure 5. Figure 5 is the sandwich test device. The cover plate and the bottom plate are Q235 steel with a radius of 70 mm, the thickness of the cover plate is 10 mm, and the bottom plate thickness is 20 mm. The size of aluminum foam is  $\Phi 78$  mm  $\times$  H78 mm, and the porosity of aluminum foam is 80%, 85%, and 90%. Each explosion impact test uses the same loading conditions; choose 520 g TNT suspended 250 mm above the center of the cover. The shock wave generated after the charge explodes acts on the cover plate and transmits and compresses the aluminum foam through the cover plate.

Set three strain measurement points on the bottom surface of the bottom plate. Strain gauge S1 is located at the center of the bottom surface of the bottom plate. Strain gauges S2 and S3 are any points on the circle 4.9 cm and 7.7 cm away from the measurement point of S1. The deformation of the bottom plate at 4.9 cm and 7.7 cm from the measuring point of the center S1 can explore the relationship between the distance and the precursor wave of the explosion. The arrangement of strain gauges is shown in Figure 6.

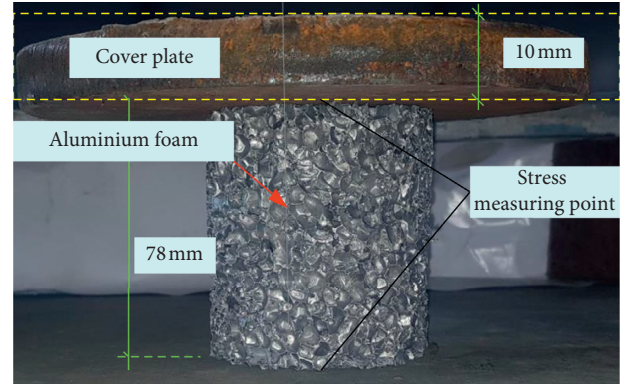


FIGURE 5: Sandwich plate antiexplosion test device.

As shown in Figure 7, the strain response time-history curve test results of the bottom plate under three porosities are shown. It can be seen from the figure that the strain waveform of the bottom plate under the three porosities has similar regular characteristics, and the strain reaches the maximum value in the first cycle of the response, and the duration of the first response cycle is about 2.5 ms. After the first cycle, the deformation of the bottom plate gradually decreases; the peak strain value from the S1 to S3 points decreases with the increase of the propagation distance. With the increase of porosity, the maximum peak value of the strain at each measurement point has increased. The increase of strain of 80%–85% is lower than 85%–90%, indicating that the porosity increases to a certain level, due to the explosion load. The fluid-solid coupling effect with the sandwich structure is enhanced, and the energy absorption effect of the aluminum foam is increased, resulting in increased strain.

Figure 8 shows the simulation results of the maximum effective strain time-history curve at each measuring point on the bottom surface of the bottom plate under three porosities. From the figure, the simulation results of the strain time-history curve are consistent with the test results. Under the three porosities, the time interval for the wave head to reach each aluminum foam plate increases slowly. The wave speed of the elastic precursor decreases with the increase of the propagation distance. The strain value at the S1 measuring point is the largest, and the vertical distance from the burst center increases. Largely, the strain value of the measuring points of S2 and S3 keeps decreasing, and the waveforms of the three measuring points are similar to S1. As the porosity increases, the strain value at each measuring point increases, indicating that as the porosity increases, the aluminum foam effectively shares the explosive load and plays the role of antiexplosion and wave elimination, which reduces the deformation of the bottom plate. Further research on the design of explosive vessels provides a theoretical basis.

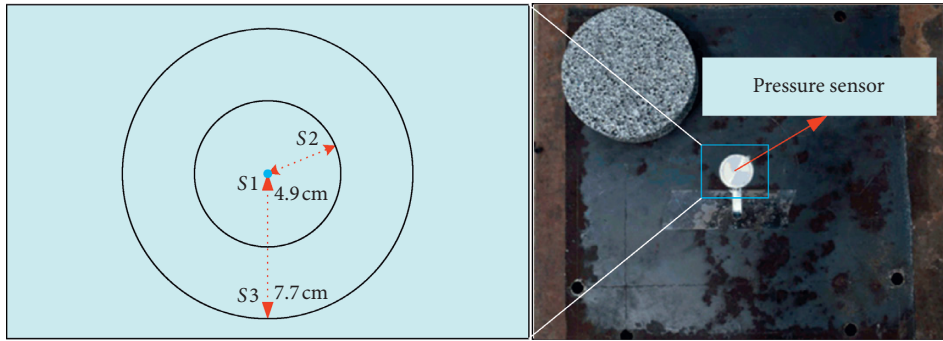


FIGURE 6: Sensor layout.

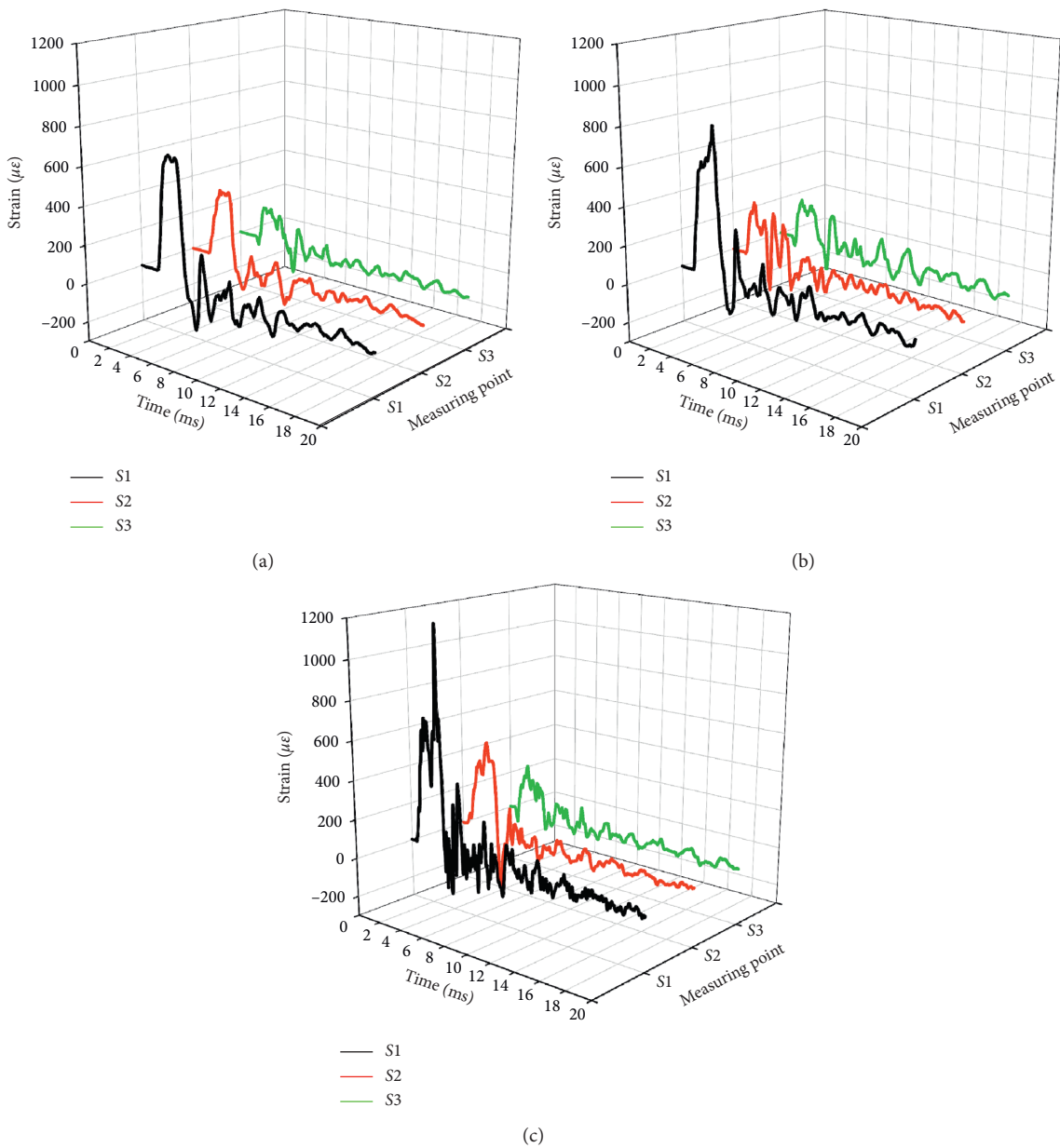


FIGURE 7: Time-history curve of strain response of bottom plate under different porosity under test. (a) 80%. (b) 85%. (c) 90%.

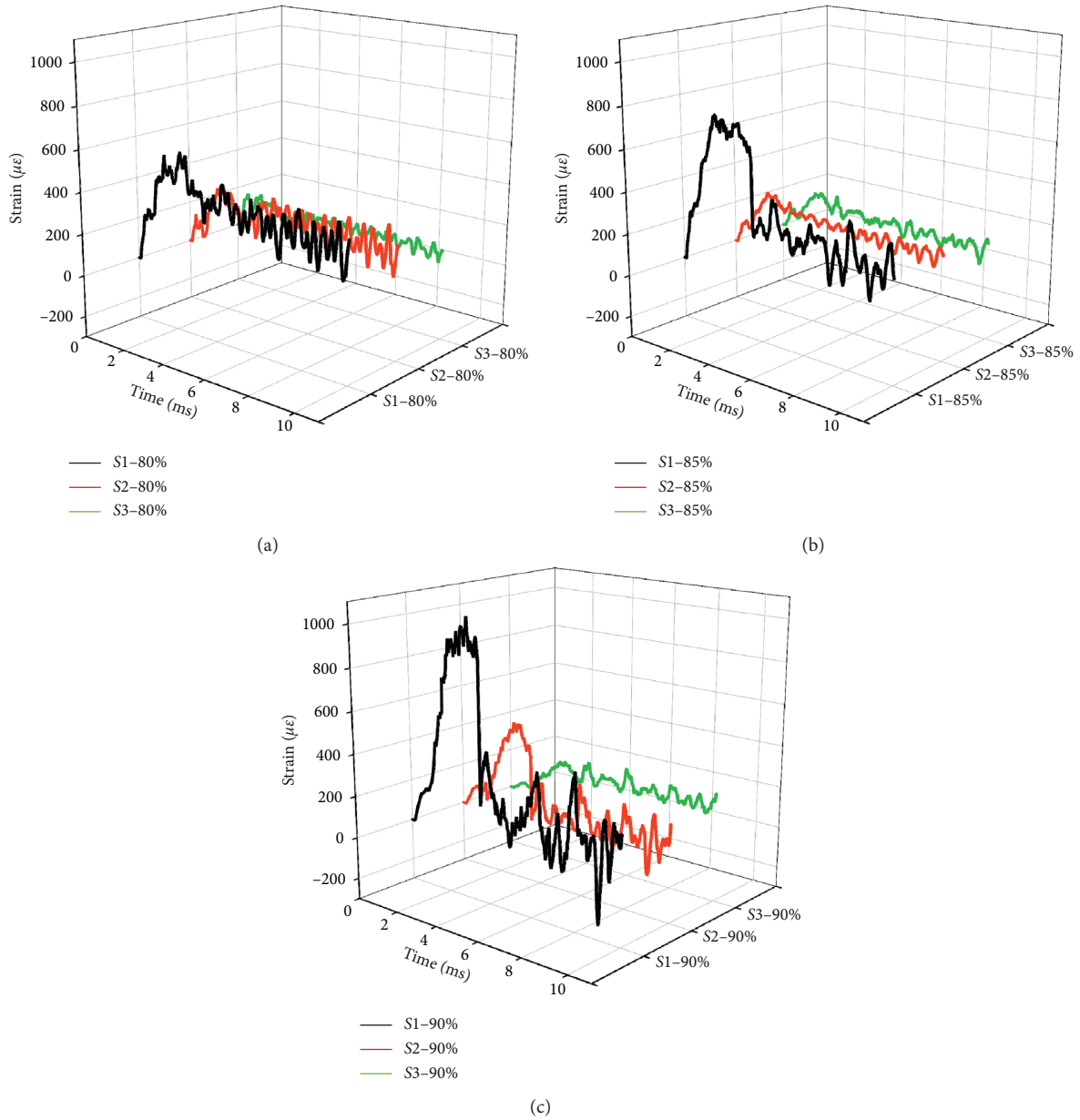


FIGURE 8: Strain time-history curve of each measuring point under different porosity under simulation. (a) 80%. (b) 85%. (c) 90%.

Figure 9 is a comparison chart of the numerical simulation results and test results of the peak strain values of the three groups of porosity aluminum foam under the explosive shock wave of the charge. In order to quantitatively describe the difference between the test and the numerical simulation, the numerical simulation results of the strain at each measuring point of the bottom plate are taken as the  $x$ -axis, and the test results are taken as the  $y$ -axis, and Figure 9 is drawn. The line with a slope of 1 in the figure indicates that the numerical simulation results are completely consistent with the test results. When the test point falls below the straight line, it means that the measured strain given by the numerical simulation is greater than the test value; when the test point falls above the straight line, it indicates that the peak value of strain at each measuring point calculated by

numerical simulation is smaller than the test value. It can be intuitively seen from Figure 9 that the test points basically fall between the solid line with a slope of 1 and the broken line with a slope of 0.90 and 1.1, and the error is within 10%. The numerical simulation results of the strain values at different measuring points under different porosities are in good agreement with the experimental results, indicating that the simulation results are good and verify the correctness of the three-dimensional mesoscopic model of aluminum foam.

*3.2. Comparison of Structural Forms.* Figure 10 is the internal structure of aluminum foam, Figure 10(a) is the macrostructure of aluminum foam in its natural state, Figure 10(b) is the morphology of 85% aluminum foam after electron

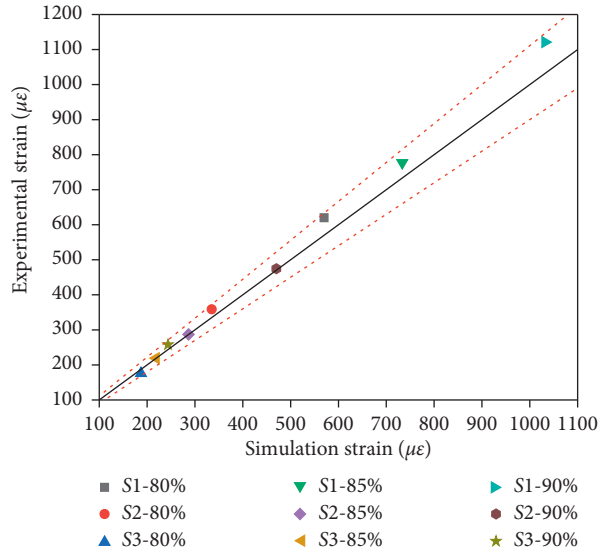


FIGURE 9: Comparison graph of numerical simulation and test results.

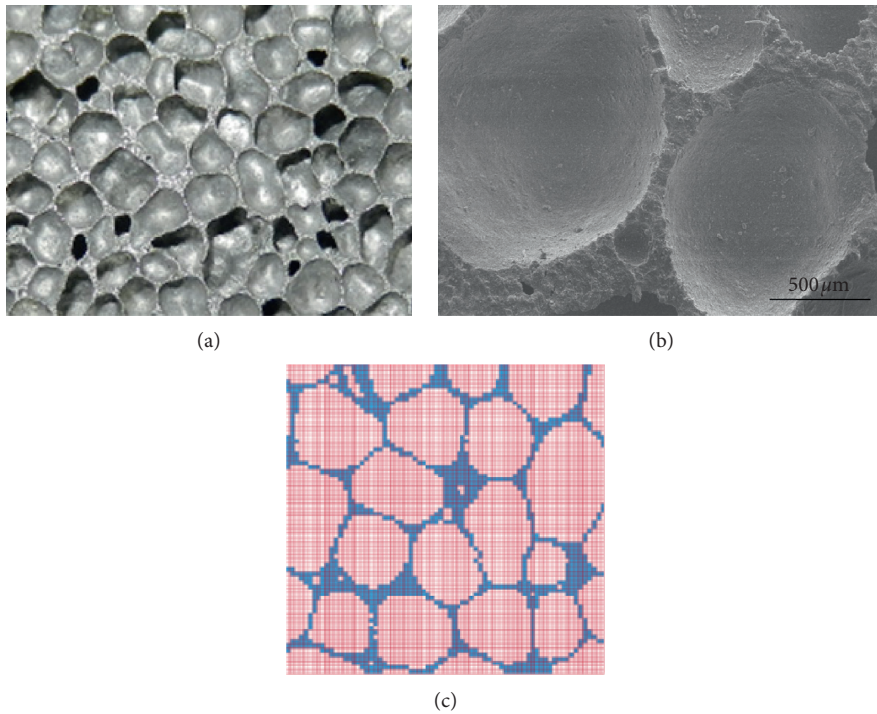


FIGURE 10: The real structure and numerical simulation structure of aluminum foam.

microscopy, and Figure 10(c) is a three-dimensional mesomodel, modeling 85% porosity aluminum foam structure. In order to verify the reliability of the structural form of the model, three slices were taken in the established three-dimensional mesomodel, the number of cell units and air units were counted, and the porosity of each slice was calculated and compared with the porosity of real aluminum foam. For comparison, the results are shown in Figure 11. It can be seen from the figure that the porosity of different slices in the

model is in good agreement with the real aluminum foam structure, with a maximum difference of 4.42%. The comprehensive strain comparison finally concluded that the agreement between the numerical simulation and the test can reach more than 91.4%. It shows that the established model can accurately express the mesoscopic structural characteristics of aluminum foam, and the modeling results are reliable. It can be observed that the structure with greater porosity has thicker cell walls.



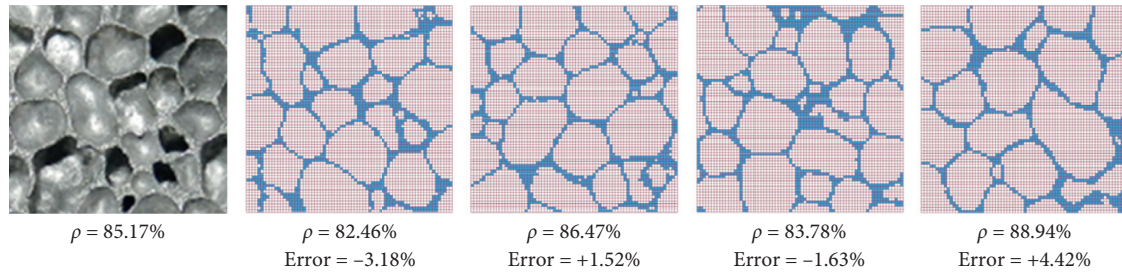


FIGURE 11: Comparison of different sections of the model with the real structure.

## 4. Analysis of Mesoresults

**4.1. Analysis of Aluminum Foam Movement Process under Blast Loading.** This section explores the movement of an aluminum foam sandwich panel with a porosity of 80% under explosive load.

In the first stage, the explosive center detonated until the shock wave reached the cover and began to compress the aluminum foam. As shown in Figure 12, at 0 ms, the explosive performs a single-point center detonation, and the shock wave front starts to expand outwards with the sphere as the center and reaches the cover plate at a certain time between 0.07 ms and 0.08 ms. At a certain speed, the aluminum foam is compressed downward, as shown in the pressure cloud diagram of the aluminum foam on the right side of Figure 12.

Figure 13 shows the process of compressing aluminum foam after the plane spherical wave reaches the cover. At 0.08 ms, the stress wave starts to propagate in the aluminum foam. At 0.09 ms, the stress wave has propagated to the bottom of the aluminum foam and transmitted the stress wave to the bottom plate, and the bottom plate has undergone a stress change. Subsequently, the foamed aluminum experienced significant plastic strain, and the cell walls began to deform and break down. At 0.429 ms, the deformation of the hole wall reached saturation, and the hole wall collapsed and destroyed. At 1.4 ms, the foam aluminum stops deforming, and the entire compression process of the aluminum foam is initially completed.

In order to explore the antiexplosion and wave attenuation capabilities of aluminum foam, we have quantitatively analyzed the attenuation effect of the stress wave strength of the aluminum foam sandwich structure through the three-dimensional mesoscopic model of the aluminum foam, taking the first layer of aluminum foam at different porosities. The average stress of the element is taken as  $\sigma_1$ , and the stress value of the foam aluminum introduced into the cover plate is obtained. The average stress of the uppermost unit of the bottom plate is taken as  $\sigma_2$ , and the stress value of the bottom plate after the explosion wave passes through the aluminum foam is obtained. The distribution of stress waves in the sandwich structure is shown in Figure 14. In the initial stage, stress concentration occurs in local weak areas, the cell walls are in an elastic state, and the stress level rises rapidly with strain. Intuitively, the stress value of 80% porosity aluminum foam in Figure 14(a) reaches 184.34 MPa at 0.4 ms, and the stress wave intensity at the bottom of the bottom plate after passing the aluminum foam is 2.45 MPa.

It can be seen that the blast shock wave decays up to 98.67% after passing through the foamed aluminum. As shown in Figure 14(b), when the porosity of the aluminum foam upper layer is 85%, the stress value reaches 147.35 MPa at 0.384 ms, and the stress wave intensity at the bottom of the bottom plate after passing the aluminum foam is 1.61 MPa. It can be seen that the blast shock wave attenuation is as high as 98.91% after passing through the aluminum foam. As shown in Figure 14(c), when the porosity of the aluminum foam upper layer is 90%, the stress value reaches 86.03 MPa at 0.149 ms, and the stress wave strength at the bottom of the bottom plate after passing the aluminum foam is 4.09 MPa. It can be seen that the explosion shock wave attenuation is as high as 95.25% after passing through the aluminum foam.

It can be concluded that as the porosity increases, the stress value of the cover plate transmitted to the aluminum foam decreases significantly, but the stress value of the bottom plate increases first as the porosity decreases.

The relationship between the foam aluminum with different porosities and the wave elimination rate is shown. As shown in Figure 15, the wave elimination ability of aluminum foam does not increase with the increase of porosity, but there is an extreme value. Probably between 83% and 84% porosity, before reaching the threshold, its antiknocking ability will increase with the increase of porosity, but after reaching the threshold, its antiknocking ability will decrease with the increase of porosity.

**4.2. Mechanism Analysis.** Figure 16 shows the effective plastic strain of the foam aluminum cell wall under the blast load. It can be seen that, at 80% porosity, the deformation is concentrated in local weak areas, and the cell walls are in an elastic state. When the porosity becomes 85%, the cell wall breaks. The cell wall deformation indicates that plastic strain consumes a lot of energy under dynamic load. As the deformation further increases, the local plastic deformation of the unit increases rapidly, and the stress reaches the bearing capacity and breaks, resulting in the generation of some weak surfaces and the formation of a single failure penetration area. As the porosity increases, the aluminum foam has stress concentration, the cell wall collapses, the plastic strain of the aluminum foam cell wall mainly occurs in the upper region, and the lower part of the aluminum foam collapses and compacts.

To explore the mechanism, we conducted an electron microscopy scan of the aluminum foam to explore its microscopic properties. Figure 17(a) shows the typical failure

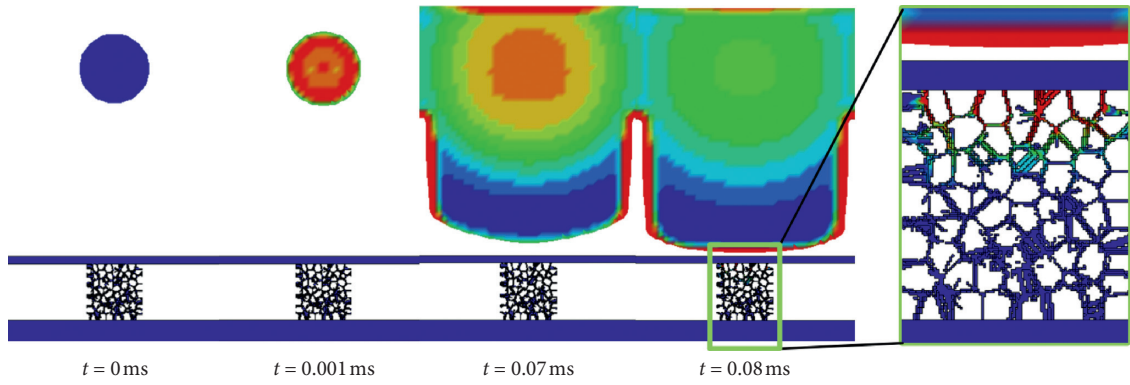


FIGURE 12: Explosion shock wave reaches the cover plate and begins to compress the aluminum foam.

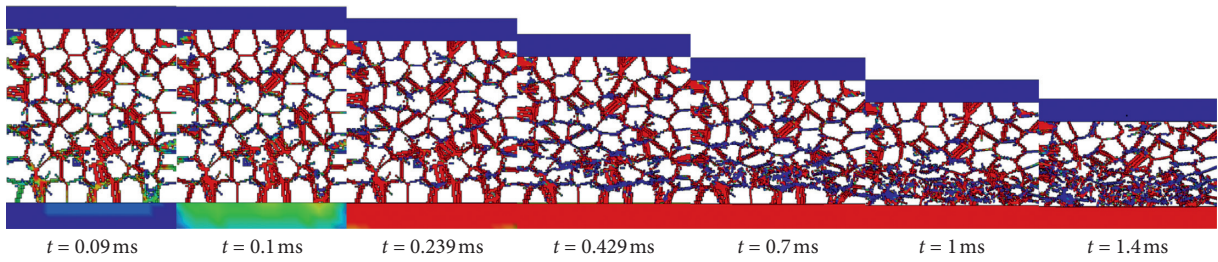


FIGURE 13: Cloud image of foam aluminum compression process.

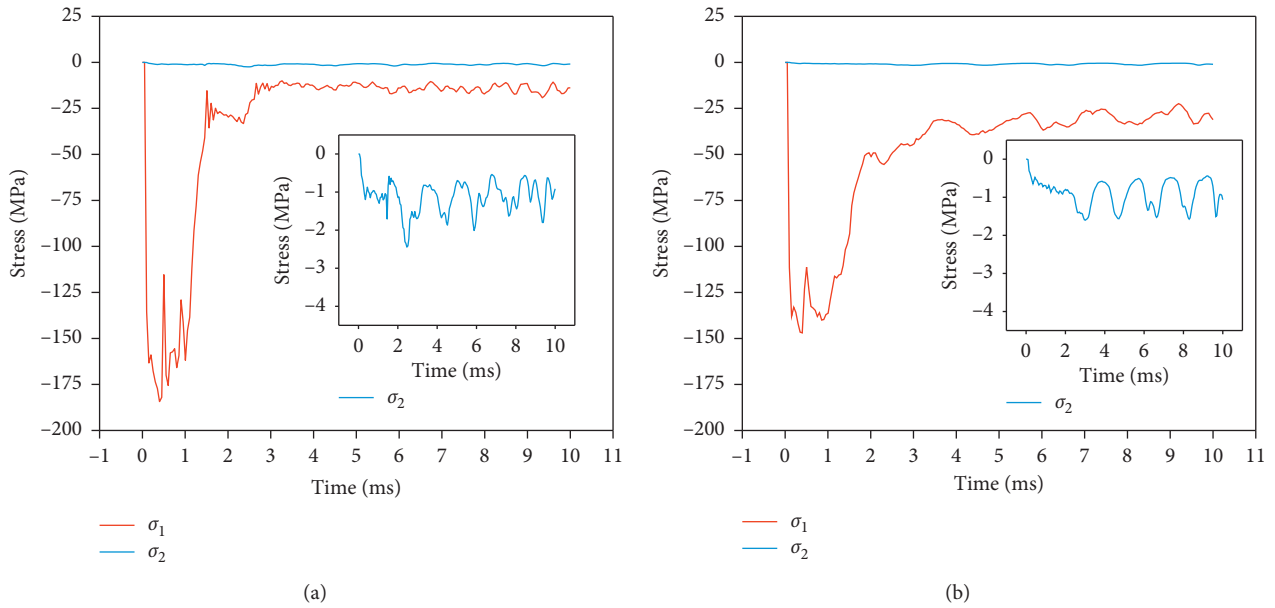


FIGURE 14: Continued.

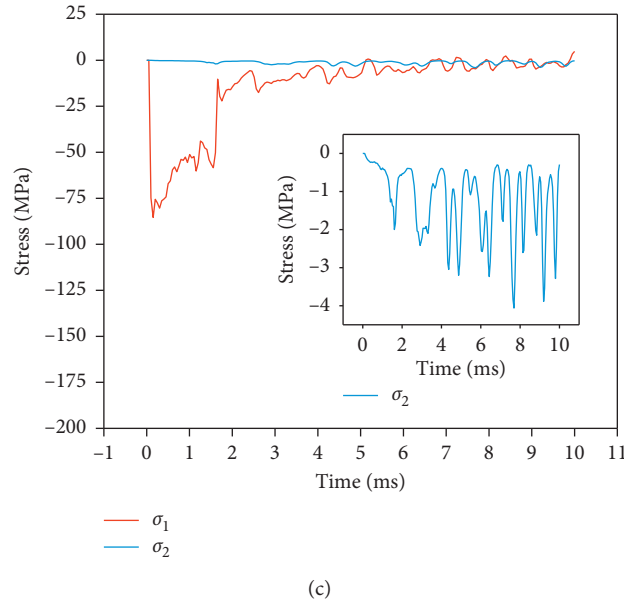


FIGURE 14: Stress-time curve of aluminum foam under different porosities. (a) 80%. (b) 85%. (c) 90%.

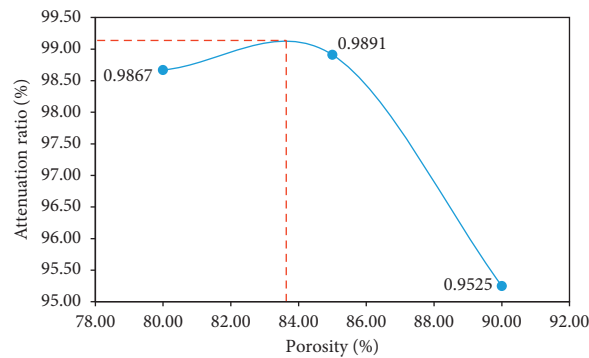


FIGURE 15: Wave elimination rate curve of aluminum foam under different porosity.

mode of aluminum foam after 85% of the sandwich structure is subjected to explosion impact. According to the results of electron microscope scanning (SEM) experiments (b-e), the microscopic morphological changes of the cells and cell walls of aluminum foam were demonstrated. Figure 16(b) shows the original state of the aluminum foam cell wall. Under the blast load, both the cover and the panel in the sandwich structure undergo large plastic deformation, and the obvious “X”-shaped failure zone appears in the aluminum foam material. According to the scanning electron microscopy results, the deformation of aluminum foam can be divided into four areas: area I, elastic deformation area, the cell structure has not changed significantly as shown in Figure 17(b); area II, the cell wall has undergone significant plastic deformation as shown in Figure 17(c); area III, the cell wall breaks and the structure is destroyed as shown in Figure 17(d); area IV, the cell wall begins to collapse and is gradually compacted as shown in Figure 17(e).

Figure 18 shows the microscopic morphology of aluminum foam under different porosity after explosion. To

ensure the single variability, the center position of aluminum foam under different porosity was selected for electron microscopy scanning. It can be seen from the figure that, with the increase of porosity, the complete pores in the aluminum foam decrease; when the porosity is 85%, the collapse of the pores in the aluminum foam is obvious, the complete pores are almost absent, and the pores are compressed. At% porosity, no complete pores can be seen inside the foam aluminum, and the degree of pore compression is significantly higher than that of the foam aluminum at 85% porosity. It can be seen that the increase in porosity will change the microstructure of aluminum foam. The greater the porosity, the deeper the degree of pore damage.

When subjected to an impact load, aluminum foam is extremely susceptible to deformation. Because aluminum foam is a porous structure, transmission and reflection occur during the propagation of an explosion wave, so there is more reflection, and a part of energy is well converted into internal energy and kinetic energy. Through the study of the mesoscopic model of aluminum foam cell walls, it is found

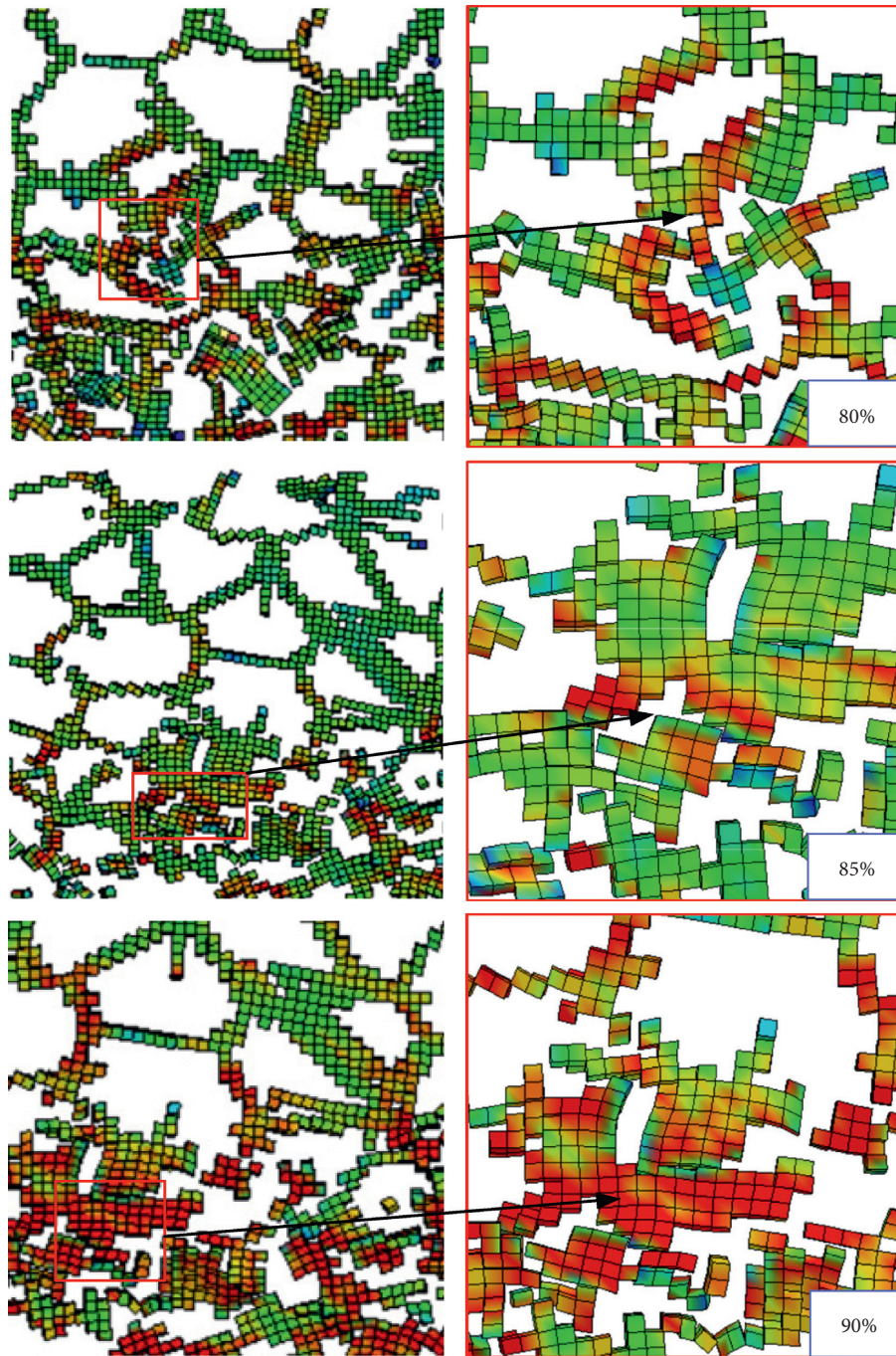


FIGURE 16: Cell wall collapse and rupture.

that the energy absorption of aluminum foam mainly consumes energy through various forms such as crushing, deformation, fracture, and mutual friction of the pore walls of the aluminum foam, so as to achieve the effect of energy absorption. When the porosity is increased to about 85%, the energy absorption effect of aluminum foam is saturated, and the porosity continues to increase so that the aluminum foam is compacted to produce stress enhancement, so the load pressure on the bottom plate increases instead. Under a certain explosion load, the explosion stress wave transmits and propagates through the solid medium of the cell wall or

between the bubbles and the solid. Due to the structural characteristics of random cell generation of aluminum foam, stress waves propagate in the constantly changing solid medium and the interference of reflection and transmission in the bubbles rapidly attenuates. When the stress wave peak is greater than the cell wall strength, deformation and destruction occur. As the propagation distance increases, the stress wave peak decreases and the amount of deformation also weakens until elastic deformation occurs. Therefore, the antiexplosion performance of foamed aluminum is directly related to the cell wall thickness and the size of bubbles.

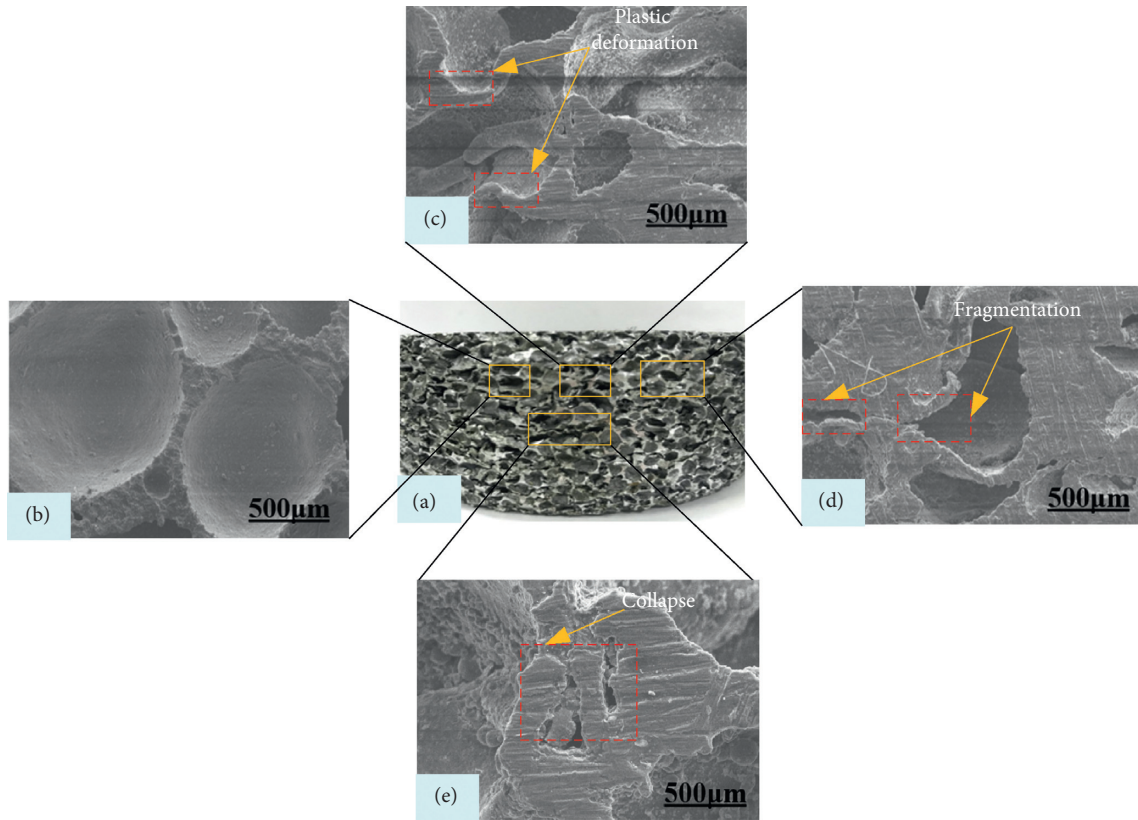


FIGURE 17: Wave elimination rate curve of aluminum foam under different porosity.

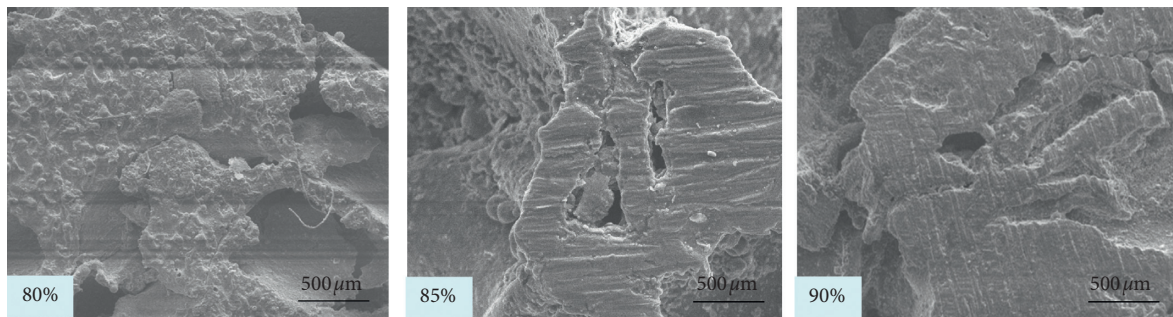


FIGURE 18: Microstructure of aluminum foam under different porosities.

Under a certain load, the contribution of the two to anti-explosion will also change. For example, after the porosity reaches the extreme value of 83%-84%, the deformation energy absorption of the cell wall is more obvious than that of the bubble. The cell wall thickness of 85% aluminum foam is greater than 90%.

## 5. Conclusion

In this paper, the explosive impact test was carried out under the explosive quantity of 520 g and the explosion height of 25 cm. The attenuation law of stress wave in the aluminum foam sandwich plate was studied, and the strain response characteristics of the bottom plate were analyzed. In

addition, a three-dimensional mesomodel for the finite element analysis of aluminum foam is established. On the basis of the three-dimensional micromechanical model, the influence of the micromechanical structure (porosity) of aluminum foam is considered, and the performance of aluminum foam under the blast impact load is studied. The following conclusions can be drawn:

- (1) Based on the Voronoi algorithm, firstly, the 2-random polyhedron model is established by using ABAQUS, and its geometric features are extracted. Then, by using Python language and Fortran language synthetically, the algorithm of generating aluminum foam with random pore size and wall thickness is written, and the grid model is established

by the algorithm read in by TrueGrid software. Finally, the part of the particles is removed to generate the aluminum foam model. Then, the energy absorption mechanism and porosity of the sandwich structure with aluminum foam core are explored under the action of explosion load, and the reliable results can be obtained, which are in good agreement with the test results.

- (2) Through the electron microscope test, it is found that the plastic strain is mainly concentrated in the upper region of aluminum foam under the action of explosion load, and there is an obvious “X” type failure zone in the material of aluminum foam. According to the SEM results, the deformation of aluminum foam is divided into four regions: elastic deformation region, plastic deformation region, cell wall fracture region, and collapse region.
- (3) Through the microscopic simulation of aluminum foam, it is found that the main reasons for aluminum foam antiknock are the deformation and destruction of the cell wall, the propagation of stress wave corresponding to the cell wall, the multiple reflections, and transmission of stress wave corresponding to bubble. Cell wall thickness and bubble size are some of the key parameters of antiknock performance.
- (4) Based on the simulation results, it is found that the porosity has a significant impact on the antiknock ability of aluminum foam, and the porosity has an extreme value of about 83% to 84% for antiknock ability. With the increase of porosity, the strain value of each measuring point of the bottom plate increases, which can absorb more energy, but it increases the energy transfer from the explosion load to the sandwich structure and makes the load amplitude acting on the bottom plate larger.

## Data Availability

The drawings and tables in the article are all original data, which are all obtained by experiments and computer simulations. In the future, when readers need the data in the article for secondary development, they can e-mail the corresponding author to provide it. The drawings and tables in the article can be edited without any problems.

## Disclosure

The authors declare that they have no financial and personal relationships with other people or organizations that can inappropriately influence their work.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The financial support from the Science and Technology Innovation Project (KYGZB0019003) and the Department

of Infrastructure Barracks (China) is gratefully acknowledged.

## References

- [1] M. F. Ashby, A. G. Evans, N. A. Fleck et al., “Metal foams: A design guide,” *Applied Mechanics Reviews*, vol. 23, no. 6, p. 119, 2012.
- [2] A. G. Evans, J. W. Hutchinson, and M. F. Ashby, “The topological design of multifunctional cellular metals,” *Progress in Materials Science*, vol. 43, p. 171, 1999.
- [3] G. W. Ma and Z. Q. Ye, “Analysis of foam claddings for blast alleviation,” *International Journal of Impact Engineering*, vol. 34, no. 1, pp. 60–70, 2007.
- [4] A. G. Evans, J. W. Hutchinson, N. A. Fleck, M. F. Ashby, and H. N. G. Wadley, “The topological design of multifunctional cellular metals,” *Progress in Materials Science*, vol. 46, no. 3–4, pp. 309–327, 2001.
- [5] J. Banhart, “Manufacture, characterisation and application of cellular metals and metal foams,” *Progress in Materials Science*, vol. 46, no. 6, pp. 559–632, 2001.
- [6] Y. Xia, C. Wu, Z. Liu et al., “Protective effect of graded density aluminium foam on RC slab under blast loading—An experimental study,” *Construction and Building Materials*, vol. 111, pp. 209–222, 2016.
- [7] S. Wang, Y. Ding, C. Wang, Z. Zheng, and J. Yu, “Dynamic material parameters of closed-cell foams under high-velocity impact,” *International Journal of Impact Engineering*, vol. 99, pp. 111–121, 2017.
- [8] N. Movahedi and S. M. H. Mirbagheri, “Comparison of the energy absorption of closed-cell aluminum foam produced by various foaming agents,” *Strength of Materials*, vol. 48, no. 3, pp. 444–449, 2016.
- [9] A. Jung, E. Lach, and S. Diebels, “New hybrid foam materials for impact protection,” *International Journal of Impact Engineering*, vol. 64, pp. 30–38, 2014.
- [10] X. R. Liu, X. G. Tian, T. J. Lu, and B. Liang, “Sandwich plates with functionally graded metallic foam cores subjected to air blast loading,” *International Journal of Mechanical Sciences*, vol. 84, pp. 61–72, 2014.
- [11] R. Sriram, U. K. Vaidya, and J. E. Kim, “Blast impact response of aluminum foam sandwich composites,” *Journal of Materials Science*, vol. 41, no. 13, pp. 4023–4039, 2006.
- [12] M. D. Goel, P. Altenhofer, V. A. Matsagar, A. K. Gupta, C. Mundt, and S. Marburg, “Interaction of a shock wave with a closed cell aluminum metal foam,” *Combustion, Explosion, and Shock Waves*, vol. 51, no. 3, pp. 373–380, 2015.
- [13] J. Shen, G. Lu, Z. Wang, and L. Zhao, “Experiments on curved sandwich panels under blast loading,” *International Journal of Impact Engineering*, vol. 37, no. 9, pp. 960–970, 2010.
- [14] L. Jing, Z. Wang, V. P. W. Shim, and L. Zhao, “An experimental study of the dynamic response of cylindrical sandwich shells with metallic foam cores subjected to blast loading,” *International Journal of Impact Engineering*, vol. 71, pp. 60–72, 2014.
- [15] Z. Wang, L. Jing, J. Ning, and L. Zhao, “The structural response of clamped sandwich beams subjected to impact loading,” *Composite Structures*, vol. 93, no. 4, pp. 1300–1308, 2011.
- [16] F. Zhu, L. Zhao, G. Lu, and Z. Wang, “Deformation and failure of blast-loaded metallic sandwich panels-experimental investigations,” *International Journal of Impact Engineering*, vol. 35, no. 8, pp. 937–951, 2008.

- [17] Y. Feng, Z. Zhu, and S. Hu, "Dynamic compressive behavior of aluminum alloy foams," *Journal of Materials Science Letters*, vol. 20, no. 18, pp. 1667-1668, 2001.
- [18] Y. An, C. Wen, P. D. Hodgson, and C. Yang, "Investigation of cell shape effect on the mechanical behaviour of open-cell metal foams," *Computational Materials Science*, vol. 55, pp. 1-9, 2012.
- [19] K. Li, X.-L. Gao, and G. Subhash, "Effects of cell shape and strut cross-sectional area variations on the elastic properties of three-dimensional open-cell foams," *Journal of the Mechanics and Physics of Solids*, vol. 54, no. 4, pp. 783-806, 2006.
- [20] S. Santosa and T. Wierzbicki, "On the modeling of crush behavior of a closed-cell aluminum foam structure," *Journal of the Mechanics and Physics of Solids*, vol. 46, no. 4, pp. 645-669, 1998.
- [21] C. Chen, T. J. Lu, and N. A. Fleck, "Effect of inclusions and holes on the stiffness and strength of honeycombs," *International Journal of Mechanical Sciences*, vol. 43, no. 2, pp. 487-504, 2001.
- [22] C. Y. Zhang, L. Q. Tang, B. Yang, L. Zhang, X. Q. Huang, and D. N. Fang, "Meso-mechanical study of collapse and fracture behaviors of closed-cell metallic foams," *Computational Materials Science*, vol. 79, pp. 45-51, 2013.
- [23] D. Karagiozova, G. N. Nurick, and G. S. Langdon, "Behaviour of sandwich panels subject to intense air blasts-Part 2: Numerical simulation," *Composite Structures*, vol. 91, no. 4, pp. 442-450, 2009.
- [24] L. J. Gibson and M. F. Ashby, *Cellular Solids: Structure and Properties*, Cambridge University Press, Cambridge, U. K, 1999.
- [25] K. Li, X.-L. Gao, and A. K. Roy, "Micromechanics model for three-dimensional open-cell foams using a tetrakaidecahedral unit cell and Castigliano's second theorem," *Composites Science and Technology*, vol. 63, no. 12, pp. 1769-1781, 2003.
- [26] M. Laroussi, K. Sab, and A. Alaoui, "Foam mechanics: Nonlinear response of an elastic 3D-periodic microstructure," *International Journal of Solids and Structures*, vol. 39, no. 13-14, pp. 3599-3623, 2002.
- [27] N. J. Mills, "The high strain mechanical response of the wet Kelvin model for open-cell foams," *International Journal of Solids and Structures*, vol. 44, no. 1, pp. 51-65, 2007.
- [28] K. Tunvir, A. Kim, and S. S. Cheon, "Analytical solution for crushing behavior of closed cell al-alloy foam," *Mechanics of Advanced Materials and Structures*, vol. 14, no. 5, pp. 321-327, 2007.
- [29] L. Li, P. Xue, Y. Chen, and H. S. U. Butt, "Insight into cell size effects on quasi-static and dynamic compressive properties of 3D foams," *Materials Science and Engineering: A*, vol. 636, pp. 60-69, 2015.
- [30] G. W. Ma, Z. Q. Ye, and Z. S. Shao, "Modeling loading rate effect on crushing stress of metallic cellular materials," *International Journal of Impact Engineering*, vol. 36, no. 6, pp. 775-782, 2009.
- [31] Z. Zheng, C. Wang, J. Yu, S. R. Reid, and J. J. Harrigan, "Dynamic stress-strain states for metal foams using a 3D cellular model," *Journal of the Mechanics and Physics of Solids*, vol. 72, pp. 93-114, 2014.
- [32] Q. Fang, J. Zhang, Y. Zhang, H. Wu, and Z. Gong, "A 3D mesoscopic model for the closed-cell metallic foams subjected to static and dynamic loadings," *International Journal of Impact Engineering*, vol. 82, pp. 103-112, 2015.
- [33] Q. Fang, J. Zhang, Y. Zhang et al., "Mesoscopic investigation of closed-cell aluminum foams on energy absorption capability under impact," *Composite Structures*, vol. 124, pp. 409-420, 2015.
- [34] D. Weaire, J. P. Kemrode, and J. Wejchert, "On the distribution of cell areas in a Voronoi network," *Philosophical Magazine B*, vol. 53, no. 5, pp. L101-L105, 1986.
- [35] D. Shi, Y. Li, and S. Zhang, *Explicit Dynamic Analysis Based on ANSYS/LS-DYNA*, Tsinghua University Press, Beijing, China, 2005.
- [36] V. R. Feldgun, Y. S. Karinski, I. Edri, and D. Z. Yankelevsky, "Prediction of the quasi-static pressure in confined and partially confined explosions and its application to blast response simulation of flexible structures," *International Journal of Impact Engineering*, vol. 90, no. 15, pp. 46-60, 2016.
- [37] H.-W. Song, Q.-J. He, J.-J. Xie, and A. Tobota, "Fracture mechanisms and size effects of brittle metallic foams: in situ compression tests inside SEM," *Composites Science and Technology*, vol. 68, no. 12, pp. 2441-2450, 2008.

## Research Article

# Study on Global Parameters Optimization of Dual-Drive Powertrain System of Pure Electric Vehicle Based on Multiple Condition Computer Simulation

Yong Wang,<sup>1,2</sup> Hongguo Cai ,<sup>3</sup> Yinghua Liao,<sup>2</sup> and Jun Gao<sup>1</sup>

<sup>1</sup>Chong Qing College of Electronic Engineering, Department of Intelligent Manufacturing and Automotive, Chongqing 401331, China

<sup>2</sup>Sichuan Provincial Key Lab of Process Equipment and Control, Sichuan University of Science & Engineering, Zigong 643002, China

<sup>3</sup>Guangxi College of Education, Department of Mathematics and Computer Science, Nanning 530023, China

Correspondence should be addressed to Hongguo Cai; [webminning@163.com](mailto:webminning@163.com)

Received 15 May 2020; Accepted 17 June 2020; Published 25 July 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Yong Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Equipped with two power sources, the dual-driving powertrain system for pure electric vehicles has a driving mode different from traditional electric vehicles. Under the premise that the structural form of the transmission system remains unchanged, the following transmission schemes can be adopted for double drive electric vehicles according to the demand power: the main and auxiliary electric transmission scheme (two motors are driven separately with dual-motor coupling drive), the transmission scheme in which the two motors always maintain coupling drive, and the speed-regulating type electric transmission scheme (the main motor is always responsible for driving, and the auxiliary motor is responsible for speed regulation). Therefore, a significant difference exists in the design methods of the power transmission system of double drive electric vehicles and existing vehicles. As for such differences, this paper adopts intelligent algorithm to design the parameters of the transmission system and introduces the genetic algorithm into the optimization design of parameters to obtain the optimal vital parameters of the power transmission system based on computer simulation. The prototype car used in this paper is a self-owned brand car; MATLAB/Simulink platform is used to build the vehicle simulation model, which is used for the computer simulation analysis of the vehicle dynamic performance and economy. It can be seen from the analysis result that the system parameters obtained by using the global optimization method proposed in this study can improve the vehicle dynamic performance and economic performance to varying degrees, which proves the efficiency and feasibility of the optimization method.

## 1. Introduction

The rapid development and popularization of pure electric vehicles cannot be realized due to their high price, imperfect charging facilities, and low driving mileage. Although the problems of high cost and incomplete charging facilities have been improved to some extent due to the relevant policies of the government and the electric vehicle technology has also achieved significant development, the critical problem which hinders the development of pure electric vehicles, namely, “how to further improve the vehicle driving mileage,” still exists. The driving mileage depends on

battery technology (to improve battery power density and energy density); also, it is closely related to other factors, such as the configuration, parameter matching, and control method of the power transmission system [1–3]. Therefore, the key to improving the driving mileage and overall performance of pure electric vehicles lie in the efficient transmission system and its design method and the perfect energy control method [4–6].

Professor Zhao adopted genetic algorithm to conduct the optimization design by computer simulation of parameters for the power transmission system of a pure electric vehicle. The dynamic performance index was used as a constraint,



and the weighting coefficient was used to transform the multiobjective optimization of driving mileage and vehicle mass into a single-objective optimization problem. The results were compared with the results obtained by the traditional optimization method. It could be seen from the comparison results that the optimization method could achieve the optimization design in an effective way [7–11]. Professor Zhou adopted two matching optimization methods (minimal kerb mass and global optimization methods) to optimize the power system parameters of the target vehicle and conducted the simulation comparison and verification of the vehicle dynamic performance and economic performance for the optimization matching effects. It could be seen from the results that the global optimization method was a more effective method, which could fully reflect the potential of parameter optimization matching of the power system in improving the vehicle performance [12]. Sornioti et al. and other scholars used the configuration of single-motor pure electric vehicle proposed by them as the basis to carry out the optimal matching calculation of transmission system parameters. They designed the parameters for two-gear transmission and compared them with those of continuously variable transmission vehicles. The results proved that the optimal matching scheme was feasible and effective [13]. Professor Zhu and other scholars adopted the method of orthogonal optimization design based on a cycle condition to carry out the optimizing calculation for the two-speed transmission ratio of pure electric vehicles. Through the simulation results, it could be seen that the dynamic performance could be significantly improved by the transmission system parameters optimized and matched by this method. Still, its economic performance had declined [14]. Professor Kowal and other scholars and Professor Spichartz and other scholars had established the optimization model for the parameter matching of power transmission system of electric vehicles and the control model for the dynamic response property of the power system, which have achieved excellent results [15, 16].

In this research, the new configuration of dual-driving powertrain system for pure electric vehicles is taken as the research goal. Based on the analysis of its working characteristics and working mode, the vehicle performance computer simulation platform is established, and the preliminary matching of power transmission system parameters is completed by computer simulation, to prove whether the initial matching is reasonable; to improve the vehicle performance, the multiobjective optimization function (including the dynamic index and economic index) is determined in this paper, and the genetic algorithm is used to optimize system parameters based on computer simulation.

## 2. Overall Structure and Working Principle of the Dual-Driving Powertrain System for Pure Electric Vehicles

The design goal of this paper is a new configuration of dual-driving powertrain system for pure electric vehicles based on

the planetary gear mechanism power coupling function. The engagement elements such as clutch and brake are used to achieve perfect working mode and the flexible switching between working modes, as shown in Figure 1.

The system consists of the following parts: one set of a single-row planetary gear mechanism, two motors (motor A; motor B), two wet brakes (B1; B2), and a wet clutch C; it outputs power through the planetary gear coupling device. The sun gear  $s$  of the single-row planetary gear mechanism is directly connected to the output shaft of motor B; the gear ring  $r$  is connected to the output shaft of motor A through the transmission gear  $d$ ; the planet carrier  $c$  is connected to the driving axle through gear  $e$  for power output; the gear ring  $r$  can be connected to the sun gear  $s$  through clutch C; the movable part of brake B1 is connected to the box, and the fixed part of brake B1 is fixed to the output shaft of motor A; the gear ring  $c$  can be braked through brake B1. The movable part of brake B2 is connected to the box, the fixed part of brake B2 is fixed to the output shaft of motor B, and the sun gear  $s$  can be braked by brake B2 [17].

The dual-driving powertrain system for pure electric vehicles belongs to the category of multipower source. In addition to being used as a mechanical power source to drive the vehicle, motor A and motor B can also be used as electric power sources, which can operate as a generator during braking. When the parking brake is required, all motors are closed, all brakes operate, and the clutch is disengaged, thus complying with the parking requirements of the vehicle; when the vehicle needs to be kept in neutral, double motors are closed, and then the brake and clutch are all disengaged; during the starting stage of the vehicle and when the vehicle is running at low- and medium-speed and small-load demands, motor A is driven separately, brake B2 operates, the sun gear is locked, and the power of motor A is input through the transmission gear and output through the planet carrier, thus realizing the small-speed ratio deceleration output and complying with the low and medium torque demands of the vehicle; when the vehicle is running at low-speed and large-load demands, motor B is driven separately, brake B1 operates, the transmission gear is locked, thus the gear ring is locked, and the power of motor B is input through the sun gear and then output through the planet carrier, thus realizing the large-speed ratio deceleration output and complying with the large torque demands of the vehicle; when the vehicle is running at high- and medium-speed and medium- and small-load demands, all brakes and clutches are disengaged, the planetary gear mechanism is unlocked, and the planetary mechanism is used to achieve the speed coupling between motor A and motor B, thus jointly driving the vehicle and realizing the continuously variable speed; when the vehicle is running at medium-speed and large-load demands, all brakes are disengaged, the clutch is engaged, the planetary gear mechanism is locked to form a entirety, and the torque coupling between motor A and motor B is realized to jointly drive the vehicle; during vehicle deceleration, the status of both brakes is controlled to determine whether the regenerative braking is performed by motor A or motor B. During the whole mode switching period, the motor can ensure that the sun gear or gear ring is

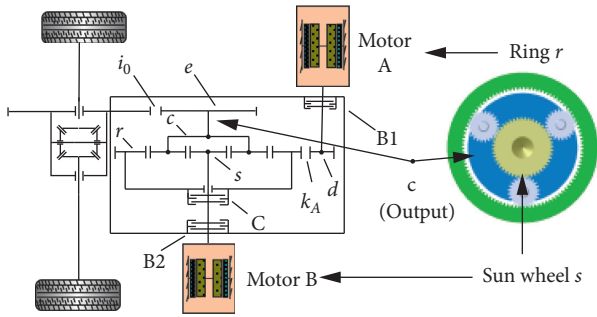


FIGURE 1: The diagrammatic sketch of dual-drive powertrain for pure electric vehicle.

subject to small resultant moment through zero-speed torque regulation and control. The motor regulation is fully utilized to realize the flexible mode switching, which is convenient for the brake and clutch disengagement or engagement and the reduction of the shift impact.

### 3. Parameter Matching Target and Scheme of Dual-Driving Powertrain System for Pure Electric Vehicles

The parameter design principles for the dual-driving powertrain system for pure electric vehicles are as follows: based on the parameters and the design performance indexes of the given vehicle model and under the premise of complying with the dynamic performance and driving mileage, the key parameters of the selected configuration and the performance indexes of the vehicle model studied in this paper are determined. See Table 1 for the vehicle performance index and vehicle parameters of the prototype vehicle.

After ensuring that the basic dynamic performance and economic performance (driving mileage) are met, this paper takes dynamic performance and economic performance as the dual objectives to carry out parameter optimization design, so as to select the key components of the power transmission system and determine the key characteristic parameters, including the drive system, the power battery system, and the power coupling system.

In this paper, the global optimization scheme is adopted, and the vehicle dynamic performance and energy consumption are taken as the dual optimization objectives. The basic dynamic performance and economic performance and the dynamic coupler characteristics are taken as the constraints. The optimization simulation platform integrating the vehicle numerical model and intelligent optimization algorithm is established. The key characteristic parameters of the power transmission system are used as optimization variables for comprehensive optimization. The key characteristic parameters of the power transmission system are optimized globally, and the optimization scheme is shown in Figure 2. The optimization scheme gives full play to the potential of the parameter matching of the power transmission system in the improvement of the vehicle performance to the greatest extent, as well as the potential of the new configuration of the dual-driving powertrain system for pure electric vehicles in energy saving [18, 19].

The impacts of various driving conditions on optimization effects are fully considered during the whole optimization process. If the optimization is based on a single driving condition, in general, the optimization results represent the optimal solution under that condition. However, if the vehicle is used for other working conditions, its adaptability may become worse and nonoptimal solution may appear. The multicondition parameter optimization method is used in this scheme, which has strong adaptability and fully considers the impacts of the efficiency characteristics of each component on the vehicle performance. In this research, intelligent optimization algorithm and continuous iterative calculation are used. Vehicle numerical model is used as the basis of performance evaluation. The optimization direction is guided, and the optimal parameter solution satisfying the requirements is searched through gradual iterations, thus achieving the global optimization design for the power transmission system parameters [20, 21].

### 4. Global Parameter Optimization of the Dual-Driving Powertrain System for Pure Electric Vehicles Based on Multiple Conditions

The parameters of the dual-driving powertrain system for pure electric vehicles are related to and affect each other. In this paper, the intelligent optimization algorithm is used to coordinate various impacts, and finally the most suitable global solution set is obtained. The parameter optimization design of dual-driving powertrain system for pure electric vehicles mainly includes the following processes: the determination of optimization objective function, the selection of optimization variables, the equation of constraint conditions, and the design of optimization algorithm.

*4.1. Efficiency Analysis of Dual-Driving Powertrain System Optimization and Processing of the Objective Function.* In this paper, the working characteristics of pure electric vehicles and the particularity of evaluation indexes are taken as research objectives; the vehicle dynamic performance and economic performance are considered. The acceleration time and driving mileage of the vehicle are taken as objective functions; the double-objective optimization function of dynamic performance and economic performance is established. The driving mileage of pure electric vehicles is improved to the largest extent on the premise of meeting the dynamic performance demand of the vehicle. In this paper, the acceleration time in 100 km is selected as the goal of dynamic performance optimization, and the driving mileage is selected as the goal of vehicle economic optimization [22].

*4.1.1. Dynamic Performance Objective Function.* The acceleration time in 100 km of the vehicle is shown as follows:  $fd_1$

TABLE 1: The performance indicators and parameter of pure electric vehicle.

|                              | Item  | Indicator  | Unit           |
|------------------------------|---|------------|----------------|
| Maximum speed                | Maximum speed                                 | $\geq 150$ | km/h           |
| Acceleration                 | 0~100 km/h acceleration time                  | $\leq 15$  | s              |
| Capability                   | 0~50 km/h acceleration time                   | $\leq 6$   | s              |
| Climbing ability             | Maximum gradient                              | $\geq 35$  | %              |
|                              | Maximum climbing speed                        | $\geq 15$  | km/h           |
| Endurance capacity           | Pure electric driving range (60 km/h)         | $\geq 150$ | km             |
|                              | Curb weight                                   | 1680       | kg             |
|                              | Windward area                                 | 2.275      | m <sup>2</sup> |
|                              | Drag coefficient                              | 0.3164     | —              |
| Prototype vehicle parameters | Rotation mass conversion coefficient $\delta$ | 1.04       | —              |
|                              | Wheel radius (dynamic)                        | 0.308      | m              |
|                              | Rolling resistance coefficient                | 0.00995    | —              |

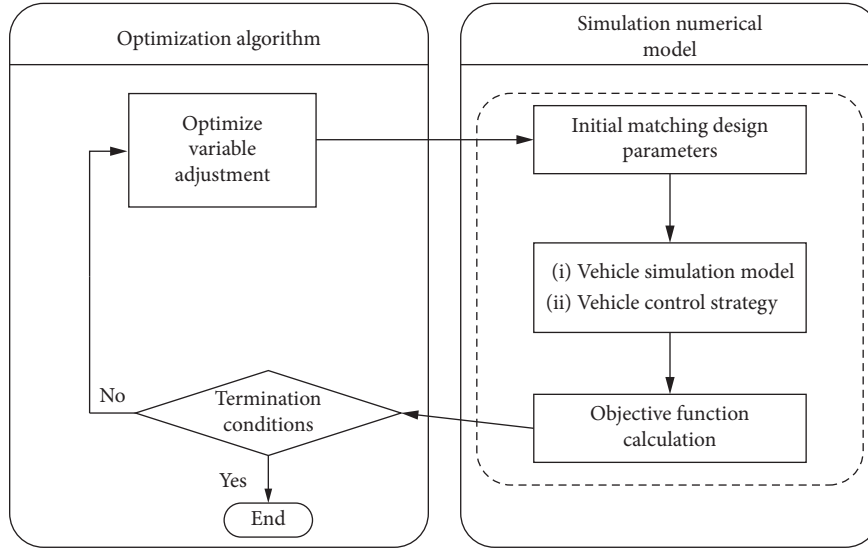


FIGURE 2: Schematic diagram of global optimization scheme.

$$t = \frac{1}{3.6} \int_0^{100} \frac{\delta m}{F_d - (mgf_r + (C_D A v^2 / 21.15))} dv, \quad (1)$$

where  $F_d$  is the maximum driving force of the vehicle.

When the vehicle speed is lower than that corresponding to the rated speed, the motor will achieve constant output at the maximum torque; when the vehicle speed is higher than that corresponding to the rated speed, the motor can achieve constant operation at the maximum power; the output torque is a function of the speed. Therefore, the maximum driving force of the vehicle is a subsection function of the vehicle speed corresponding to the rated speed, as shown in the following equation:

$$F_d = \begin{cases} \frac{3600P_{\text{sys max}}\eta_T}{v_m}, & v \leq v_m, \\ \frac{3600P_{\text{sys max}}\eta_T}{v}, & v > v_m, \end{cases} \quad (2)$$

where  $V_m$  is the vehicle speed corresponding to the rated speed of the drive system motor, in km/h.

Therefore, the dynamic performance objective function is as follows:

$$f_1(x) = t = \frac{\delta m}{3.6} \left( \int_0^{v_m} \frac{dv}{(3600P_{\text{sys max}}\eta_T/v_j) - (mgf_r + (C_D A v^2 / 21.15))} + \frac{1}{3.6} \int_{v_m}^{100} \frac{dv}{(3600P_{\text{sys max}}\eta_T/v) - (mgf_r + (C_D A v^2 / 21.15))} \right), \quad (3)$$

where  $x$  is vector composed of optimization parameters.

**4.1.2. Economic Performance Objective Function.** The average driving mileage of the fully charged vehicle under multiple cycle conditions is taken as the economic performance evaluation index in the paper, and the driving mileage under each working condition is subject to weighted average handling. Through the analysis in [17] it can be found that the driving mileage of the vehicle under single-cycle condition can be expressed by

$$L_i = \frac{Q_b \cdot U_b \cdot \text{DOD}}{1000 \cdot W_{\text{cyclei}}} \cdot s_i, \quad i = 1, 2, \dots, N, \quad (4)$$

where  $Q_b$  is capacity of power battery,  $U_b$  is total rated voltage,  $W_{\text{cyclei}}$  is the energy consumption of the vehicle when it operates under a certain cycle condition,  $S_i$  is driving mileage of vehicle under a certain cycle condition in km, and  $N$  is the number of cycle conditions.

Since the dynamic performance objective function  $f_1(x)$  represents the minimum acceleration time, it is necessary to take the reciprocal of the weighted average driving mileage, thus ensuring that it is consistent with the optimization direction of  $f_1(x)$ . At the same time, the condition weighting coefficient  $\varepsilon_i$  is introduced, which can control the weight of each condition in the process of optimization calculation,  $\sum \varepsilon_i = 1$ . Therefore, the objective function of multicondition economic performance is as follows:

$$f_2(x) = \frac{1}{\bar{L}} = \frac{1}{\sum_{i=1}^N \varepsilon_i L_i}, \quad (5)$$

where  $\bar{L}$  is average driving mileage under multiple working conditions and  $\varepsilon_i$  is working condition weighting coefficient.

**4.1.3. Transformation of Objective Function.** The multi-objective optimization problem is as follows:

$$\begin{cases} F(x) = \min[f_1(x), f_2(x), \dots, f_i(x)], & x \in \Omega, \\ dl \leq x \leq ul, \\ A_j * x = Aeq_j, \\ B_k * x \leq Beq_k. \end{cases} \quad (6)$$

In the process of solving multiobjective optimization problems, the subobjective functions may conflict with each other, and the performance improvement of one objective may reduce the performance of the other. Generally speaking, there is no absolute optimal solution which can achieve the optimization of all subobjective functions. The double-objective optimization (dynamic performance and economic performance) problem of pure electric vehicles is transformed into a single-objective optimization problem in this paper, which helps to obtain the solution and adjust the weight of each objective [23].

To transform the double-objective optimization problem of the vehicle into a single-objective optimization problem, in this paper, the above dynamic performance objective function  $f_1(x)$  and the economic performance objective

function  $f_2(x)$  are weighted and summed up [24], and the new optimization objective function is as follows:

$$f(x) = \delta_1 f_1(x) + \delta_2 f_2(x), \quad (7)$$

where  $\delta_1, \delta_2$  are weighting factor, which is greater than 0.

The weighing factor can be expressed by

$$\delta_j = \delta_{j1} \cdot \delta_{j2}, \quad j = 1, 2, \quad (8)$$

where  $\delta_{j1}$  is the subobjective weighing factor, which is used to characterize the importance of subobjectives in the process of optimization design, which is greater than zero and  $\delta_{11} + \delta_{21} = 1$ ;  $\delta_{j2}$  is a correction weighing factor, which is used to adjust the impacts of the difference in magnitude between subobjectives on the optimization calculation, which may be gradually corrected in the iterative calculation. In general, the reciprocal of the design objective is taken as the initial value. Note that  $\delta_j$  is a real number greater than zero.

**4.2. Selection of Optimization Variables.** The key parameters of the power system closely related to the vehicle performance are selected as optimization variables in this paper, namely, motor A rated power, motor A rated speed, motor B rated power, motor B rated speed, battery capacity, characteristic parameters of planetary gear, final ratio, and transmission ratio of motor A deceleration gear. The optimization variable vectors are as follows:

$$x = [P_{0A}, \omega_{0A}, P_{0B}, \omega_{0B}, Q_b, k, i_0, k_A]. \quad (9)$$

**4.3. Constraint Condition.** When setting the initial value of optimization variables and the boundary conditions of each variable, a variety of factors need to be considered comprehensively, such as the requirements of the basic performance of the vehicle, the cost of the vehicle, and the design experience. For the transmission system parameters, the principle that the components of the same system in mechanical design have similar service life shall be taken into account. In this paper, the boundary conditions of each variable are obtained after considering all influences, as shown in Table 2.

**4.4. Optimization Algorithm.** As the core of optimization calculation, optimization algorithm has a direct impact on the optimization convergence and the reliability of optimization results. Traditional optimization algorithms mainly include gradient method and steepest descent method. Such algorithms need accurate mathematical functions, as well as huge computing resources. As a computational model, genetic algorithm is based on the natural selection theory of Darwin's biological evolution theory and the biological evolution process of Mendel's genetic mechanism. It is composed of three parts (encoding and decoding, individual fitness evaluation, and genetic operation). The calculation model searches for the optimal solution complying with certain conditions by simulating the biological evolution process in nature, which is more suitable for the

TABLE 2: Optimal variable boundary condition.

| Optimization variable                                     | Initial value | Lower limit | Upper limit |
|---|---------------|-------------|-------------|
| Rated power of motor A (kW)                               | 10            | 5           | 36          |
| Rated speed of motor A (r/min)                            | 2800          | 2500        | 4000        |
| Rated power of motor B (kW)                               | 26            | 5           | 36          |
| Rated speed of motor B (r/min)                            | 3200          | 2500        | 4000        |
| Power battery capacity $Q_b$ (Ah)                         | 60            | 50          | 70          |
| Characteristic parameters of planetary gear mechanism $k$ | 3             | 1.5         | 5           |
| Final drive ratio $i_0$                                   | 2.32          | 1           | 7           |
| Motor A reduction gear ratio $k_A$                        | 6             | 1           | 8           |

optimization of iterative calculation [25, 26]. Therefore, genetic algorithm is used for the optimization design in this paper, and the optimization process is shown in Figure 3.

After repeated commissioning, the parameters for genetic algorithm are finally set as follows: the maximum evolution algebra is 80, the population size is 100, the elite number is 10, the proportion of cross offspring is 0.6, and the probability of mutation is 0.01. In addition, the rank sorting and decentralized cross function are adopted. Constrained adaptive mutation is used to achieve good convergence.

The whole optimization model includes the vehicle simulation model based on genetic algorithm toolbox in MATLAB/Simulink platform. Since the focuses of optimization calculation are the dynamic performance and economic performance, there is no need to consider the smoothness in the transient process of mode switching. Therefore, the control strategy based on logic threshold is used as the vehicle control strategy.

## 5. Global Parameter Optimization of the Dual-Drive Powertrain System for Pure Electric Vehicles Based on Multiple Conditions

In this paper, four typical conditions (NEDC, HWFET, WLTC, and CHCQ) are selected for the optimization calculation. NEDC and WLTC are usually used for regulatory testing, including low-speed and low-load condition and high-speed and high-load condition; HWFET represents high-speed and high-load condition and uniform load condition; CHCQ condition can reflect local driving characteristics. In this paper, the simulation optimization process designed above is used for the global optimization calculation of the dual-drive powertrain system for pure electric vehicles. After nearly 50 rounds of iterations, the objective function tends to be constant and the change trend is shown in Figure 4. The optimal fitness optimization variable vectors are as follows:

$$x = [12.13, 2896, 13.86, 2479, 57.71, 2.823, 4.136, 1.482]. \quad (10)$$

In view of the actual situation of the parameter design of the power transmission system, in this paper, some parameters of the dual-drive powertrain system for pure electric vehicles are rounded up, and the parameters before and after optimization are compared in Table 3.

It is found after comparing the data before and after optimization that, compared with the parameters initially

matched, the parameters after the global optimization through genetic algorithm have great changes in the following aspects: the power distribution, rated speed, and gear ratio of transmission system of the two motors. The motor power distribution is adjusted mainly based on the driving condition and the comprehensive efficiency of the driving system. If the weight of the economic target is changed, there will be some changes in the changing amplitude; the maximum speed and the gear ratio of transmission system are mainly based on the dynamic performance index (maximum vehicle speed, etc.). If the weight of dynamic performance target is changed, there will also be corresponding changes in the changing amplitude. There are no significant changes in the rated speed and the capacity of the power battery, only with a small range of changes, but the vehicle curb mass is reduced, which causes certain impacts on the dynamic performance and economic performance.

To verify whether the parameter optimization matching method of the dual-drive powertrain system for pure electric vehicles is effective, the parameters initially matched and parameters after optimization are, respectively, applied to the vehicle performance simulation platform for simulation, and its dynamic performance and economic performance are focuses of the simulation comparison. The changes in the parameters initially matched and parameters after optimization in the four typical conditions SOC are shown in Figure 5. It can be seen that the power battery SOC after optimization at the end of driving conditions is significantly higher than before, with a relatively slow change, indicating that the economic performance of parameters after optimization is significantly improved, and it can achieve more obvious energy-saving effect under high-speed conditions, indicating that the dual-drive powertrain system can improve the efficiency of driving system during high-speed cruise.

In this paper, parameters initially matched and parameters after optimization are calculated for dynamic performance, and Table 4 shows the comparison of simulation results. It can be seen from the table that the maximum vehicle speed has been increased by 9.06%, the acceleration time in 100 km has been shortened by 3.25%, and the maximum gradient has been increased by 6.33%.

In this paper, the parameters initially matched and parameters after optimization are applied to four typical conditions (NEDC, HWFET, WLTC, and CHCQ) to calculate the energy consumption and driving mileage of single cycle. The statistics and comparison of simulation results are

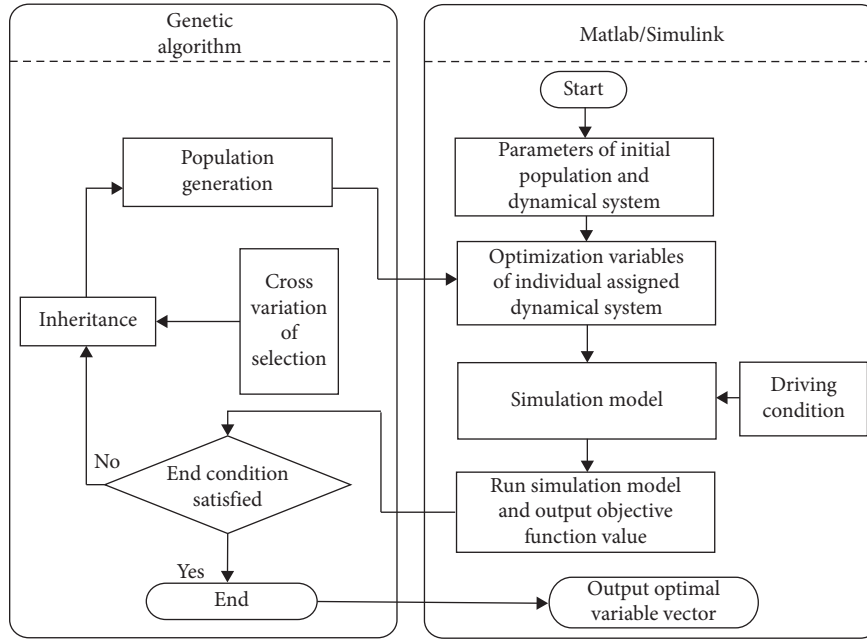


FIGURE 3: Genetic algorithms optimization flowchart.

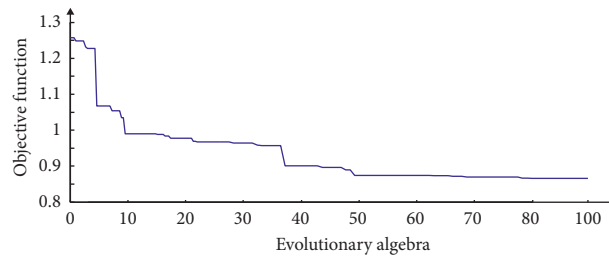


FIGURE 4: Change trend chart of objective function.

TABLE 3: Comparison of powertrain parameters after optimization.

| Parts         | Item  | Before optimization | After optimization |
|---------------|---|---------------------|--------------------|
| Motor A       | Rated/peak power (kW)                                     | 10/20               | 12/24              |
|               | Rated/maximum speed (r/min)                               | 2800/8000           | 3000/8000          |
|               | Rated/maximum torque (N·m)                                | 34/68               | 38.2/76.4          |
| Motor B       | Rated/peak power (kW)                                     | 26/52               | 24/48              |
|               | Rated/maximum speed (r/min)                               | 3200/8000           | 2500/8000          |
|               | Rated/maximum torque (N·m)                                | 77.5/155            | 91.68/183.36       |
| Powertrain    | Characteristic parameters of planetary gear mechanism $k$ | 3.8                 | 2.823              |
|               | Motor A reduction gear ratio $k_A$                        | 1.8                 | 1.482              |
|               | Final drive ratio $i_0$                                   | 3.5                 | 4.136              |
| Power battery | Rated capacity (Ah)                                       | 60                  | 58                 |

shown in Table 5. It can be seen from the table that the HWFET condition shows the most obvious effect on the economic performance improvement, with an increased range of 9.2%, which once again shows that the efficiency improvement is the most obvious in the high-speed condition. At this moment, the dual-drive powertrain system is operating in the speed-coupling mode. The economic performance simulation results prove the correctness,

feasibility, and obvious energy-saving potential of the energy-saving mechanism of the dual-drive powertrain system above.

## 6. Conclusion

The parameter matching of the power transmission system for pure electric vehicles is the premise and basis for the

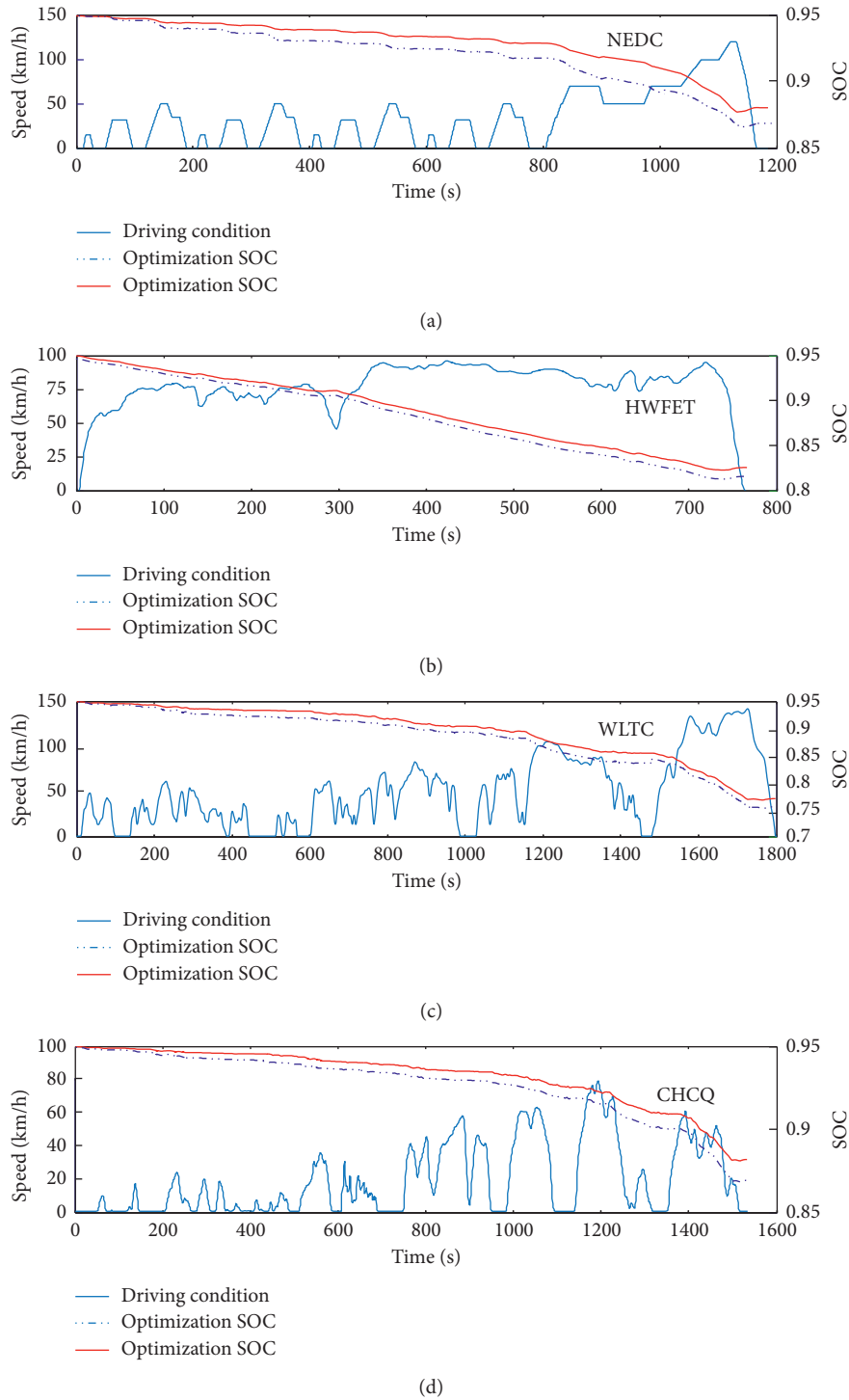


FIGURE 5: The SOC curve in different driving cycles before and after optimized parameters.

TABLE 4: Comparison of dynamic performance indicators after optimization.

| Item                          | Preliminary match | After optimization | Changes% |
|-------------------------------|-------------------|--------------------|----------|
| Maximum speed (km/h)          | 155.6             | 169.7              | 9.06     |
| 100 km acceleration time (s)  | 14.77             | 14.29              | -3.25    |
| Maximum gradient at 15 km (%) | 37.58             | 39.96              | 6.33     |

TABLE 5: Comparison of economic performance indicators after optimization.

| Cycle condition | Item                                | Preliminary match | After optimization |
|-----------------|-------------------------------------|-------------------|--------------------|
| NEDC            | Total cycle energy consumption (kJ) | 6002.22           | 5622.34            |
|                 | Driving range (km)                  | 120.76            | 128.39             |
| HWFET           | Total cycle energy consumption (kJ) | 9033.26           | 8293               |
|                 | Driving range (km)                  | 112.6             | 122.96             |
| WLTC            | Total cycle energy consumption (kJ) | 14838.4           | 13947.9            |
|                 | Driving range (km)                  | 104.8             | 111.2              |
| CHCQ            | Total cycle energy consumption (kJ) | 5324.65           | 5066.98            |
|                 | Driving range (km)                  | 127.5             | 133.67             |

follow-up research on control methods. In this paper, with the focus on the parameter matching of the dual-drive powertrain system for pure electric vehicles, the numerical model of the dual-drive powertrain system is analyzed to establish the vehicle performance simulation platform. It can be found through the simulation results that the simulation platform can simulate the working process and performance of the vehicle in an accurate manner, which lays the foundation for the performance simulation of dual-drive powertrain system for pure electric vehicles and follow-up research on control methods, to improve the vehicle performance, a multiobjective optimization function (including dynamic performance and economic performance indexes) is established in this paper, and genetic algorithm is used for the optimization design of system parameters; then the parameters before and after optimization are calculated on the vehicle simulation platform. The calculation results show that both the vehicle dynamic performance and economic performance are improved in different degrees, proving that the optimization method is efficient and feasible.

### Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Acknowledgments

The authors acknowledge the Scientific and Technological Research Program of Chongqing Municipal Education Commission (Grant no. KJQN201803106), Sichuan Provincial Key Lab of Process Equipment and Control (Grant no. GK201817), and Beijing Institute of Innovations Assistant Foundation (Grant no. 2018A05020), major projects driven by innovation in “BAGUI Scholar” Program of Guangxi Zhuang Autonomous Region of China (Grant no. 201979), and Scientific Research Project of Guangxi Education Department (Grant nos. 2020KY80007 and AA18118047-6).

### References

- [1] M. Duoba, H. Ng, and R. Larsen, “In-situ mapping and analysis of the toyota prius HEV engine,” in *Proceedings of the Future Transportation Technology Conference & Exposition*, SAE, Costa Mesa, CA, USA, August 2000.
- [2] A. Kawahashi, “A new-generation hybrid electric vehicle and its implications on power electronics,” in *Proceedings of the Annual Center for Power Electronics Systems Power Electronics Seminar*, pp. 15–20, Blacksburg, VA, USA, April 2004.
- [3] K. Muta, M. Yamazaki, and J. Tokieda, “Development of new-generation hybrid system THS II—drastic improvement of power performance and fuel economy,” in *Proceedings of the SAE 2004 World Congress & Exhibition*, SAE, Detroit, MI, USA, 2004.
- [4] L. Chu, Y. Li, and Q. N. Wang, “Study on the parametric optimization for a parallel hybrid electric vehicle power train,” in *Proceedings of the Future Transportation Technology Conference & Exposition*, SAE, Costa Mesa, CA, USA, August 2000.
- [5] Y. Dai, L. Song, and S. Cui, “Development of PMSM drives for hybrid electric applications,” *IEEE Transactions on Magnetics*, vol. 43, no. 1, pp. 434–437, 2007.
- [6] J. Aksen and K. Kurani, *The Early US Market for PHEVs: Anticipating Consumer Awareness, Recharge Potential, Design Priorities and Energy Impacts*, University of California, Oakland, CA, USA, 2008.
- [7] W. Zhao, *Parameter Matching Design of Powertrain for Battery Vehicle Based on Genetic Algorithm*, Jilin University, Changchun, China, 2017.
- [8] W. Cheng, Y. Sun, G. Li, G. Jiang, and H. Liu, “Jointly network: a network based on CNN and RBM for gesture recognition,” *Neural Computing and Applications*, vol. 31, no. 1, pp. 309–323, 2019.
- [9] B. Li, Y. Sun, G. Li et al., “Gesture recognition based on modified adaptive orthogonal matching pursuit algorithm,” *Cluster Computing*, vol. 22, no. 1, pp. 503–512, 2019.
- [10] G. Li, D. Jiang, Y. Zhou, G. Jiang, J. Kong, and G. Manogaran, “Human lesion detection method based on image information and brain signal,” *IEEE Access*, vol. 7, pp. 11533–11542, 2019.
- [11] Y. He, G. Li, Y. Liao et al., “Gesture recognition based on an improved local sparse representation classification algorithm,” *Cluster Computing*, vol. 22, no. 5, pp. 10935–10946, 2019.
- [12] F. Zhou, *Research Powertrain Parameter Design and Vehicle Control Strategy for Pure Electric Vehicle*, Jilin University, Changchun, China, 2013.



- [13] A. Sorniotti, S. Subramanyan, A. Turner, C. Cavallino, F. Viotto, and S. Bertolotto, "Selection of the optimal gearbox layout for an electric vehicle," *SAE International Journal of Engines*, vol. 4, no. 1, pp. 1267–1280, 2011.
- [14] Y. Zhu, Z. Wang, G. Han, G. Zhao, and N. Yang, "Orthogonal optimal design for drive system parameter of electric vehicle based on vehicle driving cycle," *Journal of Central South University*, vol. 44, no. 2, pp. 216–221, 2013.
- [15] J. Kowal, J. B. Gerschler, C. Schper et al., "Efficient battery models for the design of EV drive trains," in *Proceedings of the Power Electronics and Motion Control Conference*, IEEE, Ohrid, Macedonia, pp. S11–S38, September 2010.
- [16] M. Spichartz, A. Steimel, and V. Staudt, "Stator-flux-oriented control with high torque dynamics in the whole speed range for electric vehicles," in *Proceedings of the Emobility Electrical Power Train*, pp. 1–6, IEEE, Leipzig, Germany, November 2010.
- [17] Y. Wang and J. Deng, "Control strategy of dual-drive powertrain system of pure electric vehicle based on real-time optimization," *International Journal of Performability Engineering*, vol. 15, no. 12, pp. 3108–3116, 2019.
- [18] D. Jiang, G. Li, Y. Sun, J. Kong, and B. Tao, "Gesture recognition based on skeletonization algorithm and CNN with ASL database," *Multimedia Tools and Applications*, vol. 78, no. 21, pp. 29953–29970, 2019.
- [19] G. Li, H. Tang, Y. Sun et al., "Hand gesture recognition based on convolution neural network," *Cluster Computing*, vol. 22, no. 2, pp. 2719–2729, 2019.
- [20] G. Li, H. Wu, G. Jiang, S. Xu, and H. Liu, "Dynamic gesture recognition in the internet of things," *IEEE Access*, vol. 7, pp. 23713–23724, 2019.
- [21] J. Qi, G. Jiang, G. Li, Y. Sun, and B. Tao, "Intelligent human-computer interaction based on surface EMG gesture recognition," *IEEE Access*, vol. 7, pp. 61378–61387, 2019.
- [22] W. Liu and G. Yin, *Hybrid Vehicle System Modeling and Control*, China Machine Press, Beijing, China, 2015.
- [23] S. Soylu, *Electric Vehicles—Modelling and Simulations*, IntechOpen, London, UK, 2011.
- [24] L. Guzzella and A. Sciarretta, *Vehicle Drive System Modeling and optimization*, China Machine Press, Beijing, China, 2015.
- [25] W. Zhao, *Parameter Matching Design of Pure Electric Vehicle Powertrain Based on Genetic Algorithm*, Jilin University, Changchun, China, 2017.
- [26] F. Zhou, *Research on Powertrain Parameter Matching and Vehicle Control Strategy of Pure Electric Vehicle*, Jilin University, Changchun, China, 2013.

## Research Article

# Optimal Pricing Decision of Fashion Apparel considering Experiential Service and Design Enhancement

Qi Xu , Lili Zhou , and Qi Chen 

Glorious Sun School of Business and Management, Donghua University, Shanghai 200051, China

Correspondence should be addressed to Qi Xu; [xuqi@dhu.edu.cn](mailto:xuqi@dhu.edu.cn)

Received 23 May 2020; Accepted 29 June 2020; Published 21 July 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Qi Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Fashionable clothing is susceptible to seasonality, fashion popularity, and other factors. The decline in the fashion level for fashion apparel will cause its market value to continuously decrease, reducing market demand and creating a backlog of apparel inventory. Under such a circumstance, the apparel retailer chooses to maintain the fashion of the goods by providing experiential services or enhancing product design capabilities. This paper focuses on the discussions on the issue of whether experience service and design efforts are complements or substitutes. The major objective is to simultaneously determine the experience service investment and the optimal selling price to maximize the total profit. First, a Cobb–Douglas utility function is used to derive a demand function that depends on the price and fashion level. Four kinds of inventory models are further established to obtain optimal pricing and inventory ordering strategies. Second, an algorithm is presented to search for the optimal solutions of the proposed model. Finally, a numerical example is provided to perform a sensitivity analysis of the key parameters and to discuss specific managerial insights. The numerical examples show that both the experiential services and the enhanced fashion design can effectively reduce the apparel company's inventory and increase profits. When the two strategies are combined, they will produce complementary or substitution effects, which depend on the deterioration rate of the fashion level of the apparel. If the deterioration rate is less than a critical value, the interaction of experiential services and design investment has a complementarity effect.

## 1. Introduction

China is a major country in apparel production. With the advent of the era of personalized consumption, consumers are increasingly demanding fashion in apparel. However, since fashion apparel is a typical seasonal product, how to embody product innovation by enhancing the fashion level of the apparel to gain more market competitive advantage has become an urgent problem faced by apparel companies. The life cycle of fashion clothing sales is short, and its fashion level largely depends on the fashion elements contained in the clothing, such as style, pattern, material technology, and decoration accessories. Generally, the higher the fashion level is and the more fashionable the clothing is, the higher the market share of the products is. When new fashion products are launched, they will be favored and sought after by many consumers who prefer fashion. However, as time goes by, consumers' preference for original apparel gradually

decreases, so the fashion level of clothing will gradually decrease with time, and the market value of the apparel is declining. This feature usually makes consumers have different attitudes for fashion at different times. That is, the fashion level of clothing will gradually decay with time, easily causing product backlog. According to a study, the apparel industry is generally threatened by high inventories. In the first half of 2016, 79 textile and apparel listed companies had a total inventory of \$81.366 billion. On the one hand, the apparel retailer is paying more and more attention to providing virtual try-on, tie-in, fashion display, and other experiential services, so as to improve consumers' value on the apparel and, thus, to improve the fashion level of the apparel and stimulate demand [1]. On the other hand, enhancing the fashion design of clothing and continuously meeting the changing needs of the market have gradually become a new requirement for the source design of fashion clothing. Therefore, experience and design do not exist in

isolation but are closely related, and they influence each other. The decay of clothing fashion over time has an important impact on consumers' perceived value, clothing demand, and profits of clothing companies. Therefore, how to improve the interactive participation of clothing "experience and design" through experience services and clothing design investment is a very interesting and practical problem.

The existing literature related to fashion apparel mainly focused on the quick-response (QR) and price reduction. As for the research on the QR of fast fashion apparel products, Sen [2] analyzed the American fashion industry and its supply chain, and they believed that the inventory of the quick-response retailers tracked the sales of stores in real time could be kept to the minimum. Choi et al. [3] pointed out that, in the fashion industry, quick-response is a timely and influential business strategy implemented by many international companies such as Zara and H&M. Choi [4] studied the QR expectations of retailers and manufacturers and analyzed the coordination strategies of wholesale pricing and price discount contracts in the quick-response fashion supply chain. Choi [5] studied the impact of the QR plans targeting inventory services on the profits of the apparel retailer, manufacturers, and fashion supply chain systems. Cachon and Swinney [6] analyzed two key features of the fashion supply chain. One is QR production capacity, and the other is enhanced product design capabilities. All of the abovementioned studies focus on the impact of QR on inventory without considering the impact of price changes on clothing inventory during the sales period. Since price reduction promotion is a commonly used method by the apparel retailer, the promotion strategy has received extensive attention from scholars. Cachon and Feldman [7] found that frequent discounts can attract more consumers considering the consumer buying behavior, and the discount strategy is often more effective than the fixed price strategy. Cosgun et al. [8] considered the multitype clothing price reduction optimization problem of the retailer and used approximate dynamic programming algorithms to study the price reduction strategy of each product. Aviv et al. [9] studied that the seller who sells seasonal fashion goods can adopt dynamic pricing and quick inventory replenishment during the sales season to enable consumers to purchase fashion products at a higher price in advance. In view of the short life cycle of fashion products and the high uncertainty of market demand, Choi [10] considered a two-stage inventory strategy and used Bayesian methods to optimize the inventory and pricing strategies of fashion products based on actual demand data in the sales season. However, a few studies have explicitly considered clothing inventory decisions from the perspective of pricing and replenishment cycles.

Fashion clothing is a short life cycle product, and its market value is similar to the characteristics of perishable products [11]. Therefore, the research on inventory management and pricing of perishable products can be referenced for the study of the fashion apparel. Wang et al. [12] considered an inventory optimization model for perishable products whose deterioration rate followed the Weibull

distribution, and the demand depends on the instantaneous inventory level. Sicilia et al. [13] constructed a perishable inventory control model, in which a time-dependent exponential function of demand is considered. Chen et al. [14] constructed a multiperiod joint pricing and inventory control model that does not allow stock shortage. Besides, some scholars studied the inventory and pricing decisions of perishable products from the perspectives of other factors, such as purchase rate, promotional efforts, and preservation investment. Maihami and Karimi [15] analyzed the influence of the repeated purchase rate of perishable products and the speed of product diffusion on the pricing and replenishment decisions of enterprises. Tsao and Sheen [16] considered a joint decision-making on inventory and pricing of perishable products under promotional efforts. Li et al. [17] considered perishable products under the constraints of fresh-keeping investment and service investment and studied the joint pricing, ordering, and fresh-keeping investment decision-making issues of nonimmediate deterioration items. Li et al. [18] studied the stochastic production inventory model of perishable products under the dual influence of price and service level and discussed the optimal joint dynamic pricing, service, and production strategies of enterprises under the influence of external random disturbances on inventory changes.

Unlike perishable goods whose quantity and quality are declining and changing with time, the market value of fashion apparel is declining over time, but the quantity is not. Tsao [19] studied the inventory control of fashion items under the condition of trade credit and partial overstocking. Considering the effects of the fashion level and reserve price on the performance of the apparel retailer, Chen et al. [11] analyzed the optimal pricing and order strategy of the apparel retailer under dynamic and fixed cases. The literatures mentioned above did not consider the impact of experiential services on the fashion level. Chen and Xu [1] studied the optimal pricing and inventory ordering strategy in the case of no discount and discount in the sales period with considering the experiential service. Among the studies mentioned above, the work of Chen and Xu [1] and Chen et al. [11] are closest to our paper. Both studies examine the inventory and pricing decisions of fashion clothes, but they all adopt a formalized demand function, which is a linear combination of price and fashion level. However, this assumption is not in line with the reality. The difference between fashion apparel and perishable goods is that the fashion level of an apparel is a quality attribute of the apparel. Consumers can visually observe its characteristics, which is an important reference for consumers to buy. Although the price of fashion clothing is also the main factor that affects consumers' purchasing behavior, different consumers have different preferences for the price and fashion level, which will lead to the practical problem that the actual demand is difficult to determine. Some scholars have used the utility theory to study the nonlinear demand function [20–24]. For example, Su [20] took product pricing and consumer waiting costs as important factors influencing consumer utility and studied the optimal pricing strategy for retailers facing different categories of consumers. Wang et al.

[21] discussed the effect of changes in quality information asymmetry on the price decisions by using the Cobb–Douglas utility function. However, the utility functions of Su [20] and Wang et al. [21] cannot reflect the characteristics of fashion clothing, that is, the consumer’s utility function is affected by the fashion level of apparel and changes dynamically with time. Besides, the purchase quantity and the disposable income of consumers are also two important influencing factors. In short, this paper comprehensively considers the influence of fashion level, sales price, purchase quantity, and disposable income on inventory decisions.

The existing literature has conducted a lot of research on the management of perishable products, but for the fashion of clothing products, there are few studies combining experiential services and enhanced fashion design. Based on the principle of consumer utility maximization, we derive a time-varying nonlinear demand function, which mainly depends on the price and fashion level and is a nonlinear combination of the two factors. We analyze four potential operating systems—a traditional system (without experiential service and design efforts), experiential service system, design efforts system, and fast fashion systems (with experiential service and design efforts). This paper focuses the discussions on the issue of whether experience service and design efforts are complements or substitutes. The major objective is to simultaneously determine the experience service investment and the optimal selling price to maximize the total profit. It is found that both the experiential services and the enhanced fashion design can effectively reduce the apparel company’s inventory and increase profits. When the two strategies are combined, they will produce complementary or substitution effects, which depend on the deterioration rate of the fashion level of the apparel. If the deterioration rate is less than a critical value, the interaction of experiential services and design investment has a complementarity effect.

The rest of the paper is structured as follows. In Section 2, the symbols and assumptions of the model are described. Section 3 constructs four inventory models. Section 3.1 studies the retailer’s profit model and optimal pricing strategy under the basic scenario (where the apparel retailer neither provides experiential services nor enhances design). Section 3.2 discusses the case where only experiential services are provided to improve the influence of the fashion level on operational decisions; Section 3.3 investigates the situation that an apparel enterprise enhances design efforts to improve the fashion level of an apparel; Section 3.4 examines the combination of providing experiential services and enhancing design innovation and analyzes whether the two strategies are complements or substitutes. Section 4 provides numerical results and sensitivity analysis. Finally, Section 5 provides conclusions.

## 2. Notation and Assumptions

*2.1. Problem Description and Notation.* We first provide a problem description. We model an apparel retailer that sells fashionable clothing products. The demand for

fashionable clothing is determined by the selling price and fashion level of clothing during the selling period. A replenishment problem of a single fashion item is considered. The inventory system evolves as follows:  $Q$  units of items arrive at the inventory system at the beginning of each cycle. The inventory level  $I(t)$  is depicted in Figure 1 and decreases due to consumers’ demand during the replenishment cycle. Fashion apparel provides experiential service to consumers. Also, fashion apparel tries to enhance design efforts to improve the fashion level of the apparel, thus to raise the demand. The selling price and experiential service are the decision variables. We seek to answer the following questions:

- (1) Are experiential service and design efforts complements or substitutes?
- (2) How to make the optimal price and experiential service for the apparel retailer to maximize his total profit during the period?

Thus, we analyze a total of four potential operational systems. A traditional system, which represents that no experiential service or design effort is offered during the sales period, an experiential service system that employs experiential service but maintains the design efforts unchanged, a design efforts system that provides design efforts to improve the fashion level of goods but does not provide the experiential service, and a fast fashion system that employs both experiential service and design efforts.

The notations used in this paper are illustrated in Table 1.

*2.2. Utility Function.* When fashion products are released, they are favored by consumers, but as new products appear over time, consumers’ preference for original fashion clothes decreases, so the fashion level of fashion clothes will gradually decrease over time. This feature usually makes consumers buy clothes at different times to obtain different utilities, that is, the utility of consumers changes dynamically with the fashion level. Referring to the work by Wang et al. [21], a Cobb–Douglas utility function is introduced to characterize the dynamics of the utility as follows:

$$U_t = A \left( \frac{D}{p} \right)^a \omega_t^b + k(Y - pD), \quad (1)$$

where  $U_t$  is the utility of consumers at time  $t$  and  $\omega_t$  is the fashion level of apparel. The parameter  $a$  ( $0 < a < 1$ ) represents the consumers’ preference for the quantities and price of fashion clothes. The parameter  $b$  ( $0 < b < 1$ ) represents the consumers’ preference for the fashion level of the apparel. The notations  $D$ ,  $p$ , and  $Y$  represent the market demand, the selling price, and the consumers’ disposable income, respectively.  $A$  and  $k$  are positive parameters. Considering that the utility of consumers changes with time, at a certain moment, consumers will choose to purchase the optimal quantity under the circumstances of a given price and fashion to maximize their utility.

The maximization of utility provides the necessary condition for the market demand:

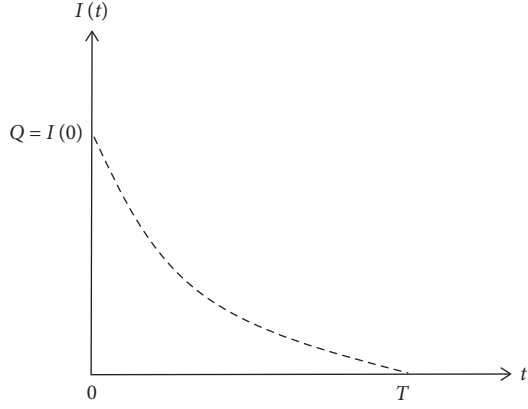


FIGURE 1: The apparel inventory level.

TABLE 1: Notations and definitions.

|            |                                 |
|------------|---------------------------------|
| $p$        | Selling price                   |
| $s$        | Experiential service investment |
| $a$        | Price preference coefficient    |
| $b$        | Fashion preference coefficient. |
| $h$        | Holding cost                    |
| $\lambda$  | Coefficient of design efforts   |
| $B$        | Fixed cost                      |
| $C_s$      | Service cost                    |
| $\eta$     | Fashion deterioration rate      |
| $D$        | Demand                          |
| $I$        | Inventory levels                |
| $Y$        | Disposable income               |
| $Q$        | Initial order quantity          |
| $c$        | Order cost per unit             |
| $U$        | Consumer utility                |
| $T$        | Replenishment cycle             |
| $\pi$      | Profit                          |
| $\omega_0$ | Initial fashion level           |
| $\omega_t$ | Fashion level                   |
| $C_\eta$   | Design cost                     |

$$D = \left(\frac{Aa}{k}\right)^{(1/1-a)} p^{-(1+a/1-a)} \omega_t^{(b/1-a)}, \quad (2)$$

where  $D$  is the market demand at time  $t$ . Substituting (2) into the utility equation (1) gives

$$U_t = \frac{k(1-a)}{a} \left(\frac{Aa}{k}\right)^{(1/1-a)} p^{-(2a/1-a)} \omega_t^{(b/1-a)} + kY. \quad (3)$$

Formula (3) shows that the rational consumer always purchases the optimal amount of fashionable clothing within a certain period of time to obtain the instantaneous maximum utility. Under the condition that other conditions remain unchanged, the maximum utility level for consumers to buy fashion clothing at different time points is different.

Next, we discuss the rationality of formula (3). First, we take the first partial derivative of (3) with respect to the price and fashion level, respectively, and we have

$$\frac{\partial U_t}{\partial p} = -2k \left(\frac{Aa}{k}\right)^{(1/1-a)} p^{-(1+a/1-a)} \omega_t^{(b/1-a)} < 0, \quad (4)$$

$$\frac{\partial U_t}{\partial \omega_t} = k \frac{a}{b} \left(\frac{Aa}{k}\right)^{(1/1-a)} p^{-(2a/1-a)} \omega_t^{(a+b/1-a)-1} > 0. \quad (5)$$

From inequalities (4) and (5), it can be observed that the utility function decreases as the sales price increases, but increases as the fashion level increases. The results given above are consistent with common sense in the industrial practice, so the assumption about the utility function of (1) is reasonable.

**2.3. Fashion Function.** It can be noted from Section 2.2 that both the consumers' utility function and the market demand for clothing are affected by the fashion level of the clothing. The complexity of this relationship arises a primitive question: how to characterize the dynamics of the fashion level? Motivated by the method used by Chen and Xu [1], this paper provides a fashion level function to describe the attenuation of the fashion level. The fashion level is given by

$$\omega_t = \omega_0 e^{-\eta t}, \quad (6)$$

where  $\omega_0$  ( $\omega_0 > 0$ ) is the initial fashion level and represents the fashion level of the garments when they were launched.  $\eta$  ( $\eta > 0$ ) is the deterioration rate of fashion level, which means the fashion level  $\omega_t$  decreases with time. We compare the influence of different initial fashion levels on the fashion level function. The fashion level decreases with time; however, the larger  $\omega_0$  is, the greater the remaining fashion level of fashion clothing at the same time. This shows that the initial fashion level has a significant delay effect on the fashion of fashion clothing. In order to meet consumers' personalized demands for fashion apparel products, retailers can provide consumers with experiential services, such as try-on, fashion display, and apparel collocation, to enhance consumers' psychological feelings about apparel fashion, so as to improve the initial fashion level of clothing to a certain extent.

Similarly, the fashion level deterioration rate  $\eta$  also has an effect on the fashion level. The smaller the  $\eta$  is, the slower the fashion level attenuation will be. A large part of the fashion level of an apparel depends on the fashion elements contained in the apparel, such as style, color, design, material technology, and decoration, and these fashion elements determine the size of the deterioration rate  $\eta$  of the fashion level of the apparel. Therefore, if retailers enhance product design and increase investment in the creative design of fashion elements through cooperation with suppliers, then the fashion level of an apparel can be reduced and relatively improved.

Substituting (6) into the demand equation (2) gives

$$D(p, t) = \left(\frac{Aa}{k}\right)^{(1/1-a)} p^{-(1+a/1-a)} \omega_0^{(b/1-a)} e^{(-b\eta/1-a)t}. \quad (7)$$

Formula (7) shows that the demand for clothing depends nonlinearly on the selling price and fashion level.

### 3. The Model

**3.1. The Traditional System.** The traditional system refers to the scenario in which the apparel retailer neither directly provides the experiential service nor indirectly invests in promoting the apparel enterprise to enhance the design during the sales period. Under the traditional scenario, the fashion level function is

$$\omega_t = \omega_0 e^{-\eta t}. \quad (8)$$

Substituting (8) into the (1) and (2), the utility function and the demand function in this section can be expressed as follows:

$$D_1 = \left(\frac{Aa}{k}\right)^{(1/1-a)} p^{-(1+a/1-a)} \omega_t^{(b/1-a)}, \quad (9)$$

$$U_1 = \frac{k(1-a)}{a} \left(\frac{Aa}{k}\right)^{(1/1-a)} p^{-(2a/1-a)} \omega_t^{(b/1-a)} + kY.$$

From Figure 1, we can see that the inventory level  $I(t)$  decreases due to consumers' demand during the time interval  $[0, T]$ . The differential equation that represents the inventory status is given by

$$\frac{dI(t)}{dt} = -D_1 = -p^{-(1+a/1-a)} \left(\frac{Aa}{k}\right)^{(1/1-a)} \omega_0^{(b/1-a)} e^{(-b\eta/1-a)t}, \quad (10)$$

With the condition  $I(T) = 0$ , solving (10) yields

$$I(t) = p^{-(1+a/1-a)} \left(\frac{Aa}{k}\right)^{(1/1-a)} \omega_0^{(b/1-a)} \cdot \left(\frac{1-a}{-b\eta}\right) (e^{(-b\eta/1-a)T} - e^{(-b\eta/1-a)t}). \quad (11)$$

The following notations are defined to simplify the expression of the mathematical model:

$$G_1 = \left(\frac{Aa}{k}\right)^{(1/1-a)} \omega_0^{(b/1-a)} \frac{1-a}{-b\eta} (e^{(-b\eta/1-a)T} - 1),$$

$$G_2 = h \left(\frac{Aa}{k}\right)^{(1/1-a)} \omega_0^{(b/1-a)} \left(\frac{1-a}{-b\eta}\right)^2 \cdot \left(\left(\frac{-b\eta T}{1-a} - 1\right) e^{(-b\eta/1-a)T} + 1\right). \quad (12)$$

The apparel retailer's ordering quantity is given as follows:

$$Q = I(0) = p^{-(1+a/1-a)} G_1. \quad (13)$$

The retailer incurs the following costs in a replenishment period: (i) purchasing cost,

$$C_1 = cQ + B = cp^{-(1+a/1-a)} G_1 + B, \quad (14)$$

where  $c$  is per unit ordering cost and  $A$  is the fixed ordering cost.

(ii) The cumulative inventory-holding cost,

$$H = h \int_0^T I(t) dt = p^{-(1+a/1-a)} G_2, \quad (15)$$

where  $h$  is unit inventory-holding cost per time unit.

In addition, the retailer's sale revenue (SR) equals to the selling price  $p$  times the sum of the demand during the replenishment period, that is,

$$SR = pQ = p^{-(2a/1-a)} G_1. \quad (16)$$

Therefore, total profit (denoted by  $\pi_1(p)$ ) during the time interval  $[0, T]$  is given by

$$\pi_1(p) = \max_{p \geq 0} \{SR - H - C_1\} = (p - c) p^{-(1+a/1-a)} G_1 - p^{-(1+a/1-a)} G_2 - B. \quad (17)$$

$\pi_1(p)$  is a function of  $p$ . Maximization of the formula (17) with respect to  $p$  yields

$$p^{(-2/1-a)+1} \left(\frac{-2a}{1-a}\right) G_1 + p^{(-2/1-a)} \left(\frac{1+a}{1-a}\right) cG_1 + p^{(-2/1-a)} \left(\frac{1+a}{1-a}\right) G_2 = 0. \quad (18)$$

By simplifying formula (17), the optimal pricing can be obtained:

$$p_1^* = \frac{(1+a)(cG_1 + G_2)}{2aG_1}. \quad (19)$$

According to the analysis mentioned above, the following proposition can be obtained.

**Proposition 1.** *When the retailer's pricing  $p$  is the optimal pricing  $p^*$ , the corresponding order quantity  $Q$  is also the optimal order quantity  $Q^*$ .*

*Proof of Proposition: 1* the retailer's optimal order quantity  $Q^*$  satisfies  $(\partial\pi_1(p)/\partial Q) = 0$ ,  $Q$  is a function of  $p$ , and we get  $\partial\pi_1(p)/\partial p = (\partial\pi_1(p)/\partial Q)\partial Q/\partial p$ . If  $p = p^*$ , there is  $\partial\pi_1(p)/\partial p|_{p^*} = 0$ , and then,  $(\partial\pi_1(p)/\partial Q)\partial Q/\partial p = 0$ ,  $\partial Q/\partial p = -(1 + a/b\eta) (Aa/k)^{(1/1-a)} p^{(-2/1-a)} \omega_0^{(b/1-a)} (1 - e^{(-b\eta T/1-a)}) < 0$ . Hence,  $\partial\pi_1(p)/\partial Q|_{Q^*} = 0$ . These steps complete the proof.  $\square$

**3.2. The Experiential Service System.** The experiential service system is based on the traditional system. Specifically, the retailer invests in experiential services, such as try-on, fashion display, and apparel collocation, in order to enhance the consumer's perception of the value of clothing fashion, thereby enhancing the consumer's psychological value. Referring to the research on fresh perishable goods preservation investment in Li et al. [17] and Dye [25], this section considers the effect of experiential service investment on improving the initial fashion level of clothing. We assume that the experiential service investment (denoted by  $s$ ) is a

continuous variable,  $s \in [0, +\infty)$ , which represents the experiential service investment of the apparel retailer in unit time. The more the investment per unit time, the greater the effort of retailers' experiential service in unit time. In order to be consistent with the law of diminishing marginal returns, it is assumed that the effect of service investment on fashion can be expressed as  $(2 - e^{-rs})\omega_0$ , that is, the initial fashion degree increases with the increase of service investment, but the incremental value is smaller and smaller. Hence,  $\lim_{s \rightarrow \infty} (2 - e^{-rs})\omega_0 = 2\omega_0$ , that is, the initial fashion degree ultimately reaches a limit value  $2\omega_0$ . In other words, we assume that the experiential service effort can double the psychological recognition of the initial fashion. Therefore, the fashion level function in an experiential service system is given by

$$\omega(t) = (2 - e^{-rs})\omega_0 e^{-\eta t}. \quad (20)$$

Experiential service means that the apparel retailer incurs the extra cost. The experiential service cost is assumed as follows:

$$C_s = \tau s^2 \int_0^T D dt, \quad (21)$$

where the positive parameter  $\tau$  is the cost coefficient of the experiential service effort. By substituting formula (20) into (1) and (2), the demand function and consumer utility function in the case of providing experiential services are obtained as follows:

$$\begin{aligned} D_2 &= \left(\frac{Aa}{k}\right)^{(1/1-a)} p^{-(1+a/1-a)} \omega_0^{(b/1-a)} (2 - e^{-rs})^{(b/1-a)} e^{-(b\eta/1-a)t}, \\ U_2 &= \frac{k(1-a)}{a} \left(\frac{Aa}{k}\right)^{(1/1-a)} p^{-(2a/1-a)} \omega_0^{(b/1-a)} \\ &\quad \cdot (2 - e^{-rs})^{(b/1-a)} e^{-(b\eta/1-a)t} + kY. \end{aligned} \quad (22)$$

According to  $(\partial U_2 / \partial s) > 0$ , the abovementioned consumer utility increases with the experiential efforts provided by the apparel retailer. Similar to that in Section 3.1, within a single cycle  $t \in [0, T]$ , we can obtain the inventory level  $I(t)$ , order quantity  $Q$ , sales revenue  $SR$ , inventory cost  $H$ , order cost  $C_1$ , and experience effort cost  $C_s$  in the case of providing experiential service:

$$\begin{aligned} I(t) &= p^{-(1+a/1-a)} \left(\frac{Aa}{k}\right)^{(1/1-a)} \omega_0^{(b/1-a)} \left(\frac{1-a}{-b\eta}\right) \\ &\quad \cdot (2 - e^{-rs})^{(b/1-a)} \left(e^{-(b\eta/1-a)T} - e^{-(b\eta/1-a)t}\right), \\ Q &= I_0 = p^{-(1+a/1-a)} (2 - e^{-rs})^{(b/1-a)} G_1, \\ SR &= pQ = p^{-(2a/1-a)} (2 - e^{-rs})^{(b/1-a)} G_1, \\ H &= h \int_0^T I(t) dt = p^{-(1+a/1-a)} (2 - e^{-rs})^{(b/1-a)} G_2, \\ C_1 &= cQ + B = cp^{-(1+a/1-a)} (2 - e^{-rs})^{(b/1-a)} G_1 + B, \\ C_s &= \tau s^2 \left[ \int_0^T D_2(p, s, t) dt \right] = \tau s^2 p^{-(1+a/1-a)} (2 - e^{-rs})^{(b/1-a)} G_1. \end{aligned} \quad (23)$$

Therefore, the total profit (denoted by  $\pi_2(p, s)$ ) during the time interval  $[0, T]$  is given by

$$\begin{aligned} \pi_2(p, s) &= \max_{p \geq 0, s \geq 0} \{SR - H - C_1 - C_s\}, \\ &= (p - c - \tau s^2) p^{-(1+a/1-a)} (2 - e^{-rs})^{(b/1-a)} G_1 \\ &\quad - p^{-(1+a/1-a)} (2 - e^{-rs})^{(b/1-a)} G_2 - B. \end{aligned} \quad (24)$$

The necessary conditions for the retailer to maximize the total profit within a single cycle are

$$\begin{aligned} \frac{\partial \pi_2(p, s)}{\partial p} &= 0, \\ \frac{\partial \pi_2(p, s)}{\partial s} &= 0. \end{aligned} \quad (25)$$

Maximization of formula (24) with respect to  $p$  yields

$$\begin{aligned} \frac{\partial \pi_2(p, s)}{\partial p} &= \left(\frac{1}{1-a}\right) p^{-(2a/1-a)} (2 - e^{-rs})^{(b/1-a)} \\ &\quad \cdot \left\{ [(1+a)(G_2 + (c + \tau s^2)G_1)] - 2aG_1 p \right\} = 0. \end{aligned} \quad (26)$$

By simplifying formula (26), the optimal pricing can be obtained as follows:

$$p_2^* = \frac{(1+a)(G_2 + (c + \tau s^2)G_1)}{2aG_1}. \quad (27)$$

Maximization of the formula (24) with respect to  $s$  yields

$$\begin{aligned} \frac{\partial \pi_2(p, s)}{\partial s} &= p^{-(1+a/1-a)} (2 - e^{-rs})^{1-(b/1-a)} \\ &\quad \cdot \left\{ \left( (p - c - \tau s^2)G_1 - G_2 \right) \frac{b}{1-a} r e^{-rs} \right. \\ &\quad \left. - 2\tau s (2 - e^{-rs})G_1 \right\} = 0. \end{aligned} \quad (28)$$

The optimal experiential service effort level  $s_2^*$  of the fashion apparel retailer satisfies the following equation:

$$bre^{-rs} \left( (p - c - \tau s^2)G_1 - G_2 \right) = 2(1-a)\tau s G_1 (2 - e^{-rs}). \quad (29)$$

Since both the equations (27) and (29) are highly nonlinear, it is difficult to obtain accurate solutions. However, Newton's iterative method can be used to find approximate solutions of equations in the real field, so we use Newton's iterative method to solve the system of simultaneous binary nonlinear equations. The specific algorithm is given in Section 3.4. By solving the model, the optimal pricing and order quantity of the fashion apparel retailer can be obtained in the experiential service system.

**3.3. The Design Innovation System.** In the era of pursuing fashion and individuation, enhancing the design of fashion clothing can improve the fashion level of products, further

making the marginal value increasing the effect of consumers on the valuation of garments. At the same time, strengthening the design of fashion apparel will also affect the supplier's production methods, which will further change the situation of coexistence and shortage of fashion products and ultimately reduce the level of apparel inventory. From the perspective of supply chain collaboration, the retailer can order products from the supplier based on the analysis of consumer purchase history data. Obviously, the initial fashion level of garments plays an important role in the retailer's ordering decision. In order to obtain fashion products with higher fashion levels, clothing retailers can provide clothing manufacturers with product design innovation subsidies, thereby improving the fashion level of clothing and reducing the decay rate of fashion levels. A smaller fashion level deterioration rate means a bigger subsidy which is charged by the manufacturer to the apparel retailer. The design innovation cost  $C_\eta$  is introduced as follows:

$$C_\eta = \lambda(\eta - \eta_1)^2 \int_0^T D dt, \quad (30)$$

where the parameter  $\lambda > 0$  represents the cost coefficient of the design innovation cost.  $\eta$  represents the deterioration rate corresponding to the standard design;  $\eta_1$  is an exogenous variable controlled by the apparel retailer. The smaller the  $\eta_1$  is, the higher the fashion level of clothing is and the more fashionable the apparel is, but the higher the cost of the design subsidy will be.

The fashion level function of the design innovation system is

$$\omega_t = \omega_0 e^{-\eta_1 t}. \quad (31)$$

Substituting (31) into the (1) and (2), the utility function and the demand function in this section can be expressed as follows:

$$D_3 = \left(\frac{Aa}{k}\right)^{(1/1-a)} p^{-(1+a/1-a)} \omega_0^{(b/1-a)} e^{(-b\eta_1/1-a)t},$$

$$U_3 = \frac{k(1-a)}{a} \left(\frac{Aa}{k}\right)^{(1/1-a)} p^{-(2a/1-a)} \omega_0^{(b/1-a)} e^{(-b\eta_1/1-a)t} + kY. \quad (32)$$

Similar to that in Section 3.1, the order quantity is

$$Q_3 = p^{-(1+a/1-a)} G_3. \quad (33)$$

Within a single cycle time  $t \in [0, T]$ , the profit function of the retailer can be obtained as follows:

$$\pi_3(p) = \max_{p \geq 0} \{SR - H - C_1 - C_\eta\},$$

$$= (p - c - \lambda(\eta - \eta_1)^2) p^{-(1+a/1-a)} G_3 - p^{-(1+a/1-a)} G_4 - B, \quad (34)$$

where  $G_3 = (Aa/k)^{(1/1-a)} \omega_0^{(b/1-a)} (1 - a - b\eta_1) (e^{(-b\eta_1/1-a)T} - 1)$  and  $G_4 = h(Aa/k)^{(1/1-a)} \omega_0^{(b/1-a)} (1 - a - b\eta_1)^2 (((-b\eta_1 T / 1 - a) - 1) e^{(-b\eta_1/1-a)T} + 1)$ .

$\pi_3(p)$  is a function of  $p$ . Maximization of the formula (34) with respect to  $p$  yields the optimal pricing:

$$p_3^* = \frac{(1+a)(cG_3 + \lambda(\eta - \eta_1)^2 G_3 + G_4)}{2aG_3}. \quad (35)$$

As can be seen from equation (35),  $p_3^*$  is affected by  $\eta_1$  of the fashion level through the enhanced design input, that is, the optimal price of the apparel retailer is affected by the deterioration rate which is determined by the apparel retailer according to the actual situation of the market.

**3.4. Fast Fashion System.** From the discussion of Sections 3.2 and 3.3, it can be observed that the apparel retailer not only provides experiential services to enhance its initial fashion but also increases design subsidy investment to reduce the consumer's psychological perception of the decay of clothing fashion, both of which can affect the operation decision of the clothing retailer. In this section, we analyze the impact of combining experiential services and design efforts in a fashion system. A fast fashion system employs both experiential services and design efforts.

The total profit of the fast fashion system is denoted by  $\pi_4$ . Specifically, we seek to answer the following question: are experiential services and design efforts complements or substitutes? If they are complements, then investing experiential services in a design innovation system results in a superadditive benefit: the incremental value of a fashion system (the change in profit over a traditional system) is more than the combined incremental value of experiential services and design efforts employed in isolation, i.e.,

$$\pi_4 - \pi_1 > (\pi_2 - \pi_1) + (\pi_3 - \pi_1). \quad (36)$$

Simplifying this expression, experiential service and design efforts are complements if and only if  $\pi_4 - \pi_2 > \pi_3 - \pi_1$ . For notational convenience, let  $\Delta\pi_{42} = \pi_4 - \pi_2$  and  $\Delta\pi_{31} = \pi_3 - \pi_1$ . If  $\Delta\pi_{42} > \Delta\pi_{31}$ , then experiential service and design efforts are complements. If  $\Delta\pi_{42} < \Delta\pi_{31}$ , then experiential service and design efforts are substitutes.

The utility function and the demand function in the fashion system can be expressed as follows:

$$D_4 = \left(\frac{Aa}{k}\right)^{(1/1-a)} p^{-(1+a/1-a)} \omega_0^{(b/1-a)} (2 - e^{-rs})^{(b/1-a)} e^{-(b\eta_1/1-a)t},$$

$$U_4 = \frac{k(1-a)}{a} \left(\frac{Aa}{k}\right)^{(1/1-a)} p^{-(2a/1-a)} \omega_0^{(b/1-a)} \cdot (2 - e^{-rs})^{(b/1-a)} e^{-(b\eta_1/1-a)t} + kY. \quad (37)$$

Considering the combined effect of experiential service and design efforts, the profit of the retailer is



$$\begin{aligned}
\pi_4(p, s) &= \max_{p \geq 0, s \geq 0} \{SR - H - C_1 - C_\eta - C_s\}, \\
&= (p - c - \tau s^2 - \lambda(\eta - \eta_1)^2) p^{-(1+a/1-a)} \\
&\quad \cdot (2 - e^{-rs})^{(b/1-a)} G_3 - p^{-(1+a/1-a)} \\
&\quad \cdot (2 - e^{-rs})^{(b/1-a)} G_4 - B.
\end{aligned} \tag{38}$$

The necessary conditions for the retailer to maximize the total profit within a single cycle are

$$\begin{aligned}
\frac{\partial \pi_4(p, s)}{\partial p} &= 0, \\
\frac{\partial \pi_4(p, s)}{\partial s} &= 0.
\end{aligned} \tag{39}$$

Maximization of formula (38) with respect to  $p$  yields

$$\begin{aligned}
\frac{\partial \pi_4(p, s)}{\partial p} &= \left( \frac{1}{1-a} \right) p^{-(2a/1-a)} (2 - e^{-rs})^{(b/1-a)} \\
&\quad \cdot \left\{ [(1+a)(G_2 + (c + \tau s^2)G_1)] - 2aG_1 p \right\} = 0.
\end{aligned} \tag{40}$$

By simplifying formula (40), the optimal pricing can be obtained as follows:

$$p_4^* = \frac{(1+a)(G_2 + (c + \tau s^2 + \lambda(\eta - \eta_1)^2)G_1)}{2aG_1}. \tag{41}$$

The optimal experiential service effort level  $s_4^*$  of the fashion apparel retailer satisfies the following equation:

$$\begin{aligned}
bre^{-rs} \left( (p - c - \tau s^2 - \lambda(\eta - \eta_1)^2) G_3 - G_4 \right) \\
= 2(1-a)\tau s (2 - e^{-rs}) G_3.
\end{aligned} \tag{42}$$

Since both the equations (41) and (42) are highly nonlinear, it is difficult to obtain accurate solutions. We use the Newton iteration method to search and solve the design algorithm of (41) and (42).

Step 1: let  $f_1 = \partial_p \pi_4(p, s) = 0$  and  $f_2 = \partial_s \pi_4(p, s) = 0$ , and the abovementioned simultaneous equations are denoted by  $\vec{f}(\vec{x}) = 0$ , where  $\vec{x} = (p, s)^T$ ,  $\vec{f}(\vec{x}) = (f_1, f_2)^T$ . The Jacobian matrix of  $\vec{f}(\vec{x})$  is denoted by  $J(\vec{x})$ .

Step 2: it is assumed that  $\vec{x}^{(i)} = (p^{(i)}, s^{(i)})^T$  is the solution to step  $i$  iteration of  $\vec{f}(\vec{x}) = 0$ ; thus,  $\vec{x}^{(i+1)} = \vec{x}^{(i)} - [J(\vec{x}^{(i)})]^{-1} \vec{f}(\vec{x}^{(i)})$ .

Step 3: it is begun with  $i = 1$  and the initial trial value of  $p^{(i)} = p^1, s^{(i)} = s^1$ .

Step 4:  $\vec{x}^{(i)} = (p^{(i)}, s^{(i)})^T$  and  $\vec{x}^{(i+1)} = (p^{(i+1)}, s^{(i+1)})^T$  are calculated.

Step 5: If  $\max\{|p^i - p^{i+1}|, |s^i - s^{i+1}|\} < 10^{-6}$ , then  $(p^*, s^*)^T = (p^{(i)}, s^{(i)})^T$  is set and proceeded to Step 6. Otherwise,  $i = i + 1$  is set and we go to Step 4.

Step 6: Using this algorithm, we obtain the optimal solution  $(p^*, s^*)$ . We can obtain and  $\pi_4^*(p, t_1)$  using that in (38).

## 4. Numerical Analysis

To illustrate the solution procedure and the results, let us apply the proposed algorithm to solve following numerical example. The results can be obtained by applying the software Matlab2016a. We set  $k = 0.5$ ,  $a = 0.35$ ,  $b = 0.65$ ,  $c = 1.5$ ,  $h = 0.5$ ,  $r = 0.5$ ,  $A = 120$ ,  $\omega_0 = 0.8$ ,  $T = 15$ ,  $Y = 10000$ ,  $\eta = 0.015$ , and  $\tau = 1$ ,  $\lambda = 2$ .

**4.1. Optimal Solutions.** Figure 2 shows that the profit  $\pi_1$  of the retailer is a strictly concave function of the selling price  $p$  in the traditional system. As shown in Figure 2, the global maximum can be obtained at the local maximum point, that is, there exists an optimal sales price. The scenario of the design innovation system is similar to that of the traditional system. As can be seen from Figure 3, the profit function  $\pi_2$  is strictly concave in  $(p, s)$ . Therefore, we ensure that a local maximum is the global maximum. The scenario of the fashion system is similar to that of the experiential system. Hence, we can conclude that the maximum searched by the algorithm could be viewed as the optimal solution of the problem.

**4.2. Impact of the Preference Combinations  $(a, b)$ .** In the case of providing experiential services, in order to better explain the model, we specifically analyze the impact of experiential service efforts and prices on retailer decisions under different combination preferences. From Table 2, it can be observed that when consumers' fashion preference for apparel  $b$  is greater than price preference  $a$ , the retailer should invest more in experiential service efforts to improve the initial fashion of the apparel. At this time, the sales price is gradually increasing. On the contrary, consumers have little preference for clothing fashion, and retailers lack the incentive to invest in service.

**4.3. Impact of the Deterioration Rate  $(\eta)$ .** Next, we will analyze the impact of the fashion level deterioration rate  $\eta_1$  on the retailer's decision-making in the case of enhancing design. Note that  $\eta_1$  is exogenous variables. Here, we adjust the value of  $\eta_1$  such that it varies from 0.007 to 0.015 at a step of 0.001 and keep the values of other parameters unchanged, as shown in Table 3. We can see from Table 4 that when  $\eta_1 = \eta = 0.015$ , the design cost function is  $C_\eta = 0$ . It represents that the apparel retailer did not invest in fashion design, which is equivalent to the traditional system. With decrease in  $\eta_1$ , that is, the increase in fashion design reduces the deterioration rate of the fashion level of an apparel, the price and order quantity are correspondingly enhanced, and the consumer's utility and retailer's profit are also increased. However, when  $\eta_1$  reaches 0.009, the order quantity of the retailer does not decrease, and the profits and consumers' utility reach an inflection point and, then, start to decrease. This is because of the decrease in fashion deterioration rate benefits from the design that enhances fashion. When it is reduced to a certain extent, the cost of the effort to enhance the design rises rapidly, and the retailer shares part of the

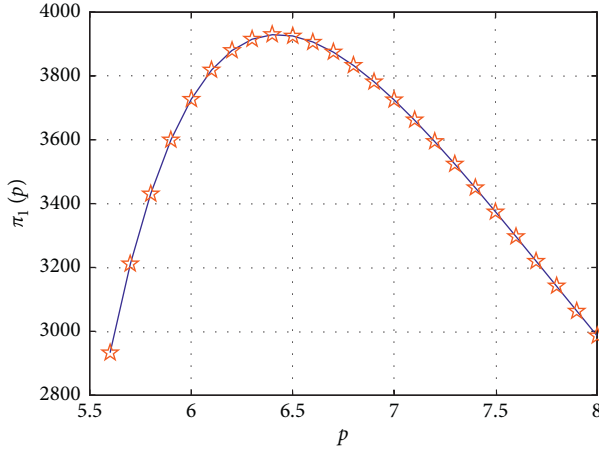


FIGURE 2: The effect of price on profits.

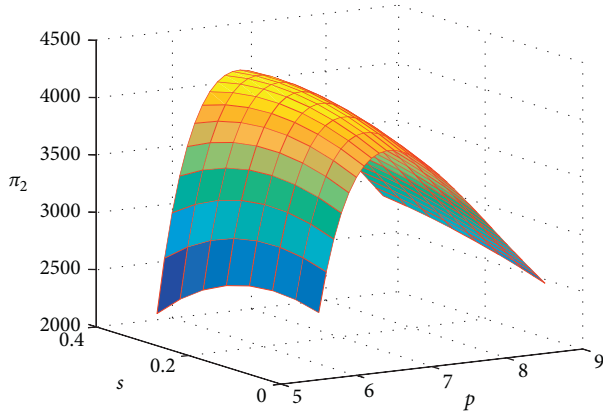


FIGURE 3: Impact of price and experiential services on profitability.

TABLE 2: Impact of different preference combinations on the decisions of the apparel retailer.

| $a$  | $b$  | $p^*$ | $s^*$ | $Q^*$   | $U^*$   | $\pi_2^*$ |
|------|------|-------|-------|---------|---------|-----------|
| 0.7  | 0.3  | 6.21  | 0.22  | 9628.48 | 5894.40 | 10427.28  |
| 0.65 | 0.35 | 6.52  | 0.27  | 3128.75 | 5376.36 | 4205.38   |
| 0.6  | 0.4  | 6.89  | 0.32  | 1316.38 | 5203.17 | 2146.34   |
| 0.55 | 0.45 | 7.65  | 0.41  | 456.02  | 5104.68 | 981.60    |
| 0.5  | 0.5  | 7.88  | 0.44  | 364.71  | 5092.65 | 838.56    |
| 0.45 | 0.55 | 8.53  | 0.44  | 224.99  | 5076.01 | 607.69    |
| 0.4  | 0.6  | 9.49  | 0.60  | 137.85  | 5060.13 | 440.46    |
| 0.35 | 0.65 | 10.71 | 0.70  | 89.06   | 5052.76 | 339.29    |

design cost through the order volume, thus offsetting part of the sales revenue. Therefore, investment in fashion design needs to be within a certain range. The optimal design investment can reduce the deterioration rate of the fashion level and increase the profit at the same time. However, excessive emphasis on the fashion level will inevitably increase the investment in fashion design and indirectly affect the income of the apparel retailer.

We have analyzed the impact of experiential service investment and design efforts on apparel retailers' operational decisions. Next, we analyze whether experiential

TABLE 3: Impact of the deterioration rate on retailer's decisions.

| $\eta_1$ | $p^*$ | $Q^*$    | $U^*$    | $\pi_3^*$ |
|----------|-------|----------|----------|-----------|
| 0.015    | 6.485 | 2949.128 | 5385.960 | 3936.917  |
| 0.014    | 6.497 | 2945.019 | 5388.597 | 3938.682  |
| 0.013    | 6.509 | 2941.004 | 5391.256 | 3940.559  |
| 0.012    | 6.521 | 2937.083 | 5393.937 | 3942.549  |
| 0.011    | 6.533 | 2933.256 | 5396.639 | 3944.654  |
| 0.010    | 6.545 | 2929.523 | 5401.960 | 3946.873  |
| 0.009    | 6.521 | 2962.014 | 5403.748 | 3949.207  |
| 0.008    | 6.533 | 2966.497 | 5398.173 | 3947.658  |
| 0.007    | 6.545 | 2967.293 | 5395.626 | 3945.225  |

TABLE 4: Impact of the deterioration rate on the optimal profit.

| $\eta_1$ | $\pi_1^*$ | $\pi_2^*$ | $\pi_3^*$ | $\pi_4^*$ |
|----------|-----------|-----------|-----------|-----------|
| 0.02     | 3936.917  | 4215.028  | 3929.758  | 4205.302  |
| 0.019    | 3936.917  | 4215.028  | 3930.970  | 4207.012  |
| 0.018    | 3936.917  | 4215.028  | 3932.292  | 4208.839  |
| 0.017    | 3936.917  | 4215.028  | 3933.723  | 4210.783  |
| 0.016    | 3936.917  | 4215.028  | 3935.264  | 4212.846  |
| 0.015    | 3936.917  | 4215.028  | 3936.917  | 4215.028  |
| 0.014    | 3936.917  | 4215.028  | 3938.682  | 4217.330  |
| 0.013    | 3936.917  | 4215.028  | 3940.559  | 4219.752  |
| 0.012    | 3936.917  | 4215.028  | 3942.549  | 4222.296  |
| 0.011    | 3936.917  | 4215.028  | 3944.654  | 4224.962  |
| 0.010    | 3936.917  | 4215.028  | 3946.873  | 4227.751  |

service and design efforts are substituted or complementary. Consider four different scenarios: the basic scenario (the case where the apparel retailer provides neither experiential services nor enhanced design), the experiential service scenario alone, the enhanced design scenario alone, and the experiential service and enhanced design combination due to the retailer's revenue are affected by the fashion level of clothing, so this paper will discuss the impact of the fashion deterioration rate on the optimal profit and, then, analyze whether the experiential services and design efforts are an alternative relationship or a complementary relationship. The sensitivity analysis of the deterioration rate results is shown in Table 4.

Table 4 shows that the profits  $\pi_2^*$ ,  $\pi_3^*$ , and  $\pi_4^*$  are always larger than the profit  $\pi_1^*$  in the traditional system. This result indicates that investing in experiential services alone, product design innovation alone, or implementing both strategies can help apparel retailers increase profits. Besides, the profit  $\pi_4^*$  of the fashion system is larger than the experiential one  $\pi_2^*$  until the deterioration rate of the fashion level reaches a certain threshold  $\hat{\eta}_1 = 0.015$ . In other words, when  $\eta_1$  is small ( $\eta_1 \in (0.01, 0.015)$ ),  $\pi_4^* > \pi_2^*$ , the retailer will further improve profits by offering design efforts on the basis of providing experiential service. Experiential service and promotion efforts are complements. When  $\eta$  is relatively large ( $\eta \in [0.015, 0.02)$ ),  $\pi_4^* < \pi_2^*$ , and this is in contrast to previous policies. Moreover, we conduct a sensitivity analysis of the deterioration rate of the fashion level, which is shown in Figure 4.

As we can see from Figure 4, when  $\eta_1$  is relatively large ( $\pi_4^* < \pi_2^*$ ),  $\Delta\pi_{42} < \Delta\pi_{31}$ . Thus, experiential services and

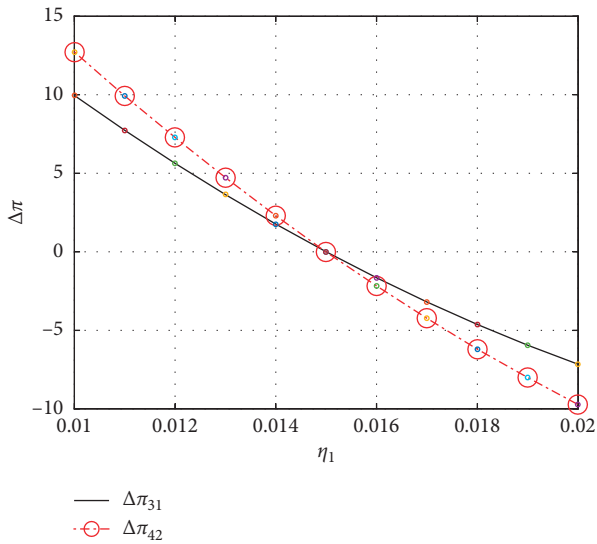


FIGURE 4: The incremental value of an interactive fashion over experiential services ( $\Delta\pi_{42}$ ) and the incremental value of design efforts over the traditional system ( $\Delta\pi_{31}$ ), plotted as a function of the deterioration rate of fashion level,  $\eta_1$ .

design efforts are substitutes. Moreover, the difference between  $\Delta\pi_{31}$  and  $\Delta\pi_{42}$  is increasing in  $\eta_1$ . When  $\eta_1$  is large (costume designing is relatively poor), the substitution effect is dramatic. This occurs because the larger the deterioration rate  $\eta_1$  is, the faster the fashion level  $\omega(t)$  fading is. If the costume's design is relatively poor, it is difficult for the apparel retailer to improve the fashion level of clothing by providing experiential services. These apparels are very outdated, resulting in backlog. Besides, when  $\eta_1$  is relatively small ( $\pi_4^* > \pi_2^*$ ),  $\Delta\pi_{42} > \Delta\pi_{31}$ . Thus, experiential services and design efforts are complementary. Moreover, the difference between  $\Delta\pi_{31}$  and  $\Delta\pi_{42}$  is decreasing in  $\eta_1$ . When  $\eta_1$  is small (the design of the clothing is very novel), the complementarity effect is dramatic. This occurs because the smaller the deterioration rate  $\eta_1$  is, the slower the fashion level  $\omega(t)$  fading is. If the design of the clothing is very novel, the clothing retailer can enhance the brand by providing experiential services, thereby attracting more consumers to purchase products and obtain more profits.

## 5. Conclusions

The decline in the fashion level of fashion for fashion apparel has caused its market value to continuously decrease, reducing market demand and creating a backlog of clothing inventory. The demand for fashionable clothing is determined by the selling price and fashion degree of clothing during the selling period. Under these backgrounds, the apparel retailer invests in experiential services, such as providing virtual and physical try-on, to improve the fashion level of apparel and, thus, to increase the demand. On the other hand, the fashion apparel adopts a design innovation strategy, that is, the apparel manufacturer improves the innovative design of the apparel according to the requirements of retailers in order to produce an apparel that meets a

certain attenuation criterion. We present a modeling framework that allows us to capture and isolate the key aspects that define a fast fashion system: experiential service and design efforts. By employing this approach, we analyze four potential operating systems—a traditional system (without experiential service and design efforts), experiential service system, design efforts system, and fast fashion systems (with experiential service and design efforts). We characterize the properties of the optimal solution (the selling price and experiential service investment) in each case and propose an iterative algorithm. This paper focuses on discussing the issue of whether experiential service and design efforts are complements or substitutes. We find that when the deterioration rate of the fashion level is relatively small, the incremental value of a fashion system (the change in profit over a traditional system) is more than the combined incremental value of experiential service and design efforts employed in isolation. In other words, the interaction of experiential service and design efforts has a complementarity effect when the design of the clothing is relatively novel.

In the proposed inventory model, the price was assumed as a static variable. As we all know, price strategy, as an important factor in the success of the business of the modern firms, has a significant effect on the consumers' demand. The static price strategy ignores the fashion attribute of fashion apparel, which has been declining over time. In fact, dynamic pricing is a powerful tool to manage the demand of different quality level products. Therefore, according to the time-varying characteristics of fashion clothing, how to implement the time-varying dynamic pricing strategy would be worth exploring in the future. Besides, this paper considers the deterministic situation that we may also consider the stochastic demand in the model. The pricing and experiential service investment policy in the supply chain of fashion items should also be considered. Last but not the least, it would be useful to test our results through qualitative research and case studies.

## Data Availability

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also form part of an ongoing study.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (71572033 and 71832001).

## References

- [1] Q. Chen and Q. Xu, "Optimal inventory and pricing decision of fashion the apparel retailer considering experiential service

- efforts,” *Chinese Journal of Management*, vol. 15, no. 10, pp. 1089–1097, 2018.
- [2] A. Şen, “The US fashion industry: a supply chain review,” *International Journal of Production Economics*, vol. 114, no. 2, pp. 571–593, 2008.
  - [3] T.-M. Choi, C.-L. Hui, N. Liu, S.-F. Ng, and Y. Yu, “Fast fashion sales forecasting with limited data and time,” *Decision Support Systems*, vol. 59, pp. 84–92, 2014.
  - [4] T.-M. Choi, “Quick response fashion supply chains in the big data Era,” in *Optimization and Control for Systems in the Big-Data Era*, pp. 253–267, Springer, Berlin, Germany, 2017.
  - [5] T.-M. Choi, “Inventory service target in quick response fashion retail supply chains,” *Service Science*, vol. 8, no. 4, pp. 406–419, 2016.
  - [6] G. P. Cachon and R. Swinney, “The value of fast fashion: quick response, enhanced design, and strategic consumer behavior,” *Management Science*, vol. 57, no. 4, pp. 778–795, 2011.
  - [7] G. P. Cachon and P. Feldman, “Price commitments with strategic consumers: why it can be optimal to discount more frequently . . . than optimal,” *Manufacturing & Service Operations Management*, vol. 17, no. 3, pp. 399–410, 2015.
  - [8] Ö. Cosgun, U. Kula, and C. Kahraman, “Markdown optimization for an apparel retailer under cross-price and initial inventory effects,” *Knowledge-Based Systems*, vol. 120, pp. 186–197, 2017.
  - [9] Y. Aviv, M. M. Wei, and F. Zhang, “Responsive pricing of fashion products: the effects of demand learning and strategic consumer behavior,” *Management Science*, vol. 65, no. 7, pp. 2982–3000, 2019.
  - [10] T.-M. Choi, “Pre-season stocking and pricing decisions for fashion retailers with multiple information updating,” *International Journal of Production Economics*, vol. 106, no. 1, pp. 146–170, 2007.
  - [11] Q. Chen, Q. Xu, and W. Wang, “Optimal policies for the pricing and replenishment of fashion apparel considering the effect of fashion level,” *Complexity*, vol. 2019, Article ID 9253605, 12 pages, 2019.
  - [12] D. Wang, L. Chen, and C. Yang, “Study on optimized inventory model of perishable products based on weibull function,” *Industrial Engineering and Management*, vol. 19, no. 5, pp. 1–6, 2014.
  - [13] J. Sicilia, M. González-De-la-Rosa, J. Febles-Acosta, and D. Alcaide-López-de-Pablo, “An inventory model for deteriorating items with shortages and time-varying demand,” *International Journal of Production Economics*, vol. 155, pp. 155–162, 2014.
  - [14] L. Chen, X. Chen, M. F. Kebliş, and G. Li, “Optimal pricing and replenishment policy for deteriorating inventory under stock-level-dependent, time-varying and price-dependent demand,” *Computers & Industrial Engineering*, vol. 135, pp. 1294–1299, 2019.
  - [15] R. Maihami and B. Karimi, “Optimizing the pricing and replenishment policy for non-instantaneous deteriorating items with stochastic demand and promotional efforts,” *Computers & Operations Research*, vol. 51, pp. 302–312, 2014.
  - [16] Y.-C. Tsao and G.-J. Sheen, “Dynamic pricing, promotion and replenishment policies for a deteriorating item under permissible delay in payments,” *Computers & Operations Research*, vol. 35, no. 11, pp. 3562–3580, 2008.
  - [17] G. Li, X. He, J. Zhou, and H. Wu, “Pricing, replenishment and preservation technology investment decisions for non-instantaneous deteriorating items,” *Omega*, vol. 84, pp. 114–126, 2019.
  - [18] S. Li, J. Zhang, and W. Tang, “Joint dynamic pricing and inventory control policy for a stochastic inventory system with perishable products,” *International Journal of Production Research*, vol. 53, no. 10, pp. 2937–2950, 2015.
  - [19] Y.-C. Tsao, “Two-phase pricing and inventory management for deteriorating and fashion goods under trade credit,” *Mathematical Methods of Operations Research*, vol. 72, no. 1, pp. 107–127, 2010.
  - [20] X. Su, “Intertemporal pricing with strategic customer behavior,” *Management Science*, vol. 53, no. 5, pp. 726–741, 2007.
  - [21] Q. Wang, S. Cai, and Y. Tang, “Study of customer utility and enterprise profit based on quality information asymmetry degree,” *Chinese Journal of Management Science*, vol. 14, no. 1, pp. 88–93, 2006.
  - [22] Z. Huang, X. Xu, J. Ni, H. Zhu, and C. Wang, “Multimodal representation learning for recommendation in Internet of Things,” *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10675–10685, 2019.
  - [23] M. Zhang, D. Zhang, H. Yao, and K. Zhang, “A probabilistic model of human error assessment for autonomous cargo ships focusing on human-autonomy collaboration,” *Safety Science*, vol. 130, p. 104838, 2020.
  - [24] J. Yang, N. Xiong, A. V. Vasilakos et al., “A fingerprint recognition scheme based on assembling invariant moments for cloud computing communications,” *IEEE Systems Journal*, vol. 5, no. 4, pp. 574–583, 2011.
  - [25] C.-Y. Dye, “The effect of preservation technology investment on a non-instantaneous deteriorating inventory model,” *Omega*, vol. 41, no. 5, pp. 872–880, 2013.

## Research Article

# Application Study of Sigmoid Regularization Method in Coke Quality Prediction

Shaohong Yan,<sup>1,2</sup> Hailong Zhao,<sup>2,3</sup> Liangxu Liu,<sup>3</sup> Qiaozhi Sang,<sup>2</sup> Peng Chen,<sup>2</sup> and Jie Li<sup>3</sup>

<sup>1</sup>College of Sciences, North China University of Science and Technology, Tangshan 063200, China

<sup>2</sup>College of Metallurgy and Energy, North China University of Science and Technology, Tangshan 063200, China

<sup>3</sup>North China University of Science and Technology Innovation of Mathematical Modeling Laboratory, Tangshan 063200, China

Correspondence should be addressed to Jie Li; [lijie@ncst.edu.cn](mailto:lijie@ncst.edu.cn)

Received 7 April 2020; Accepted 22 June 2020; Published 20 July 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Shaohong Yan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Coke is an indispensable and vital flue for blast furnace smelting, during which it plays a key role as a reducing agent, heat source, and support skeleton. Models of prediction of coke quality based on ANN are established to map the functional relationship between quality parameters  $M_b$ ,  $A_d$ ,  $V_{daf}$ ,  $S_{t,d}$ , and caking property ( $X$ ,  $Y$ , and  $G$ ) of mixed coal and quality parameters  $A_d$ ,  $S_{t,d}$ , coke reactivity index (CRI), and coke strength after reaction (CSR) of coke. A regularized network training method based on Sigmoid function is designed considering that redundancy of network structure may lead to the learning of undesired noise, in which weights having little impact on performance and leading to overfitting are removed in terms of computational complexity and training errors. The cascade forward neural network with validation is found to be the most suitable one for coke quality prediction, with errors around 5%, followed by feedforward neural network structure and radial basis neural networks. The cascade forward neural network may play a guiding role during the coke production.

## 1. Introduction

With the growing trend of large-scale blast furnace, smelting effect of blast furnace and its economic and technical indicators are more deeply influenced by the quality and performance of coke [1]. Unfortunately, various types and proportions of mixed coal can lead to different coke quality [2]. In that case, it is of significance to study the relationship between physical and chemical properties of mixed coal and that of coke in controlling the coke quality.

Due to multidimension, being dynamic, incompleteness, and uncertainty, the data of coking test is difficult to collect and handle, and this makes the actual production process still in an unscheduled and irregular state and causes the occurrence of the “rich data and poor knowledge” phenomenon [3]. Rule data mining technology is capable of finding the potential link between historical data, promoting information transmission, digging knowledge from data, and providing decision basis [4]. Numerous researches and experiments have been carried out in response to coke

quality prediction. Alvarez, R et al. predicted the CSR property of coke using the law of addition [5]. Roest et al. used different statistical analysis tools (MLR and PCR) and the ANN technique to solve the same problem [6]. Using the thought of coal blend with petrography, Tao Peisheng guided the coke quality prediction and determination of the coal blending ratio based on coal-rock phase composition and reflectance pattern of single coal while taking into account coal maceral and coke strength, etc. [7]. Although the above studies enrich the theory of coke quality prediction models, rare deep data rule mining is adopted for relevant data [8]. A large number of models predicting coke quality have been proposed up to now, a majority of which are merely based on coal characteristics and limited to the same coal geography origin, but no general applicable prediction formula has been developed up to now [9].

The statistical law of abundant coking testing data indicates the extremely strong nonlinear relationship between the physical and chemical properties of coke and that of mixed coal [10]. A variety of existing artificial coal blending

plans are inefficient and the application of corresponding results is quite poor, which fails to meet the production need of modern enterprises [11]. Fortunately, the adoption of neural network technology can decrease time consumption and reduce economic costs through rule mining of the experimental data and rationally predicting the physical and chemical properties of coke from mixed coal.

The application of data rule mining techniques is here and there right now. It becomes an important mission of interdisciplinary researchers in coking industry and mathematical application fields to guide the data mining process and draw meaningful results using the professional knowledge in coking industry.

## 2. Factors Affecting Coke Quality

In steel industry, coke as a fuel has been used to provide heat for the melting of slag and metal [12]. Coke is also used as a reduction agent to reduce iron ore to elemental iron. Hence, the parameters influencing the coke property are of great importance [5]. Coke production is shown in Figure 1. First of all, the raw materials are collected from the raw material yard to the mixing warehouse, crushed by a crusher, then put into a coke oven for calcination, and finally put into a blast furnace for smelting.

**2.1. Mixed Coal Properties.** The properties of mixed coal play key roles in determining the coke quality. Moisture, ash, volatile, sulfur, and caking property affect it in different degrees. Table 1 shows the range of content of mixed coal corresponding indicators and the impact on coke production.

**2.1.1. Full Moisture ( $M_t$ ).** It is the sum of intrinsic and external moisture in coal. Too much moisture content is bad for processing and transportation; it also deteriorates thermal stability and thermal conductivity when burning, reduces the coke yield, and extends the coking cycle when coking. Hence, it is specified to be below 8% [13].

**2.1.2. Ash ( $A_d$ ).** Ash in coal remains in coke after coking. Too much ash will lead to sudden drop of the cold state strength, the increase of residue in a blast furnace, and finally production reduction. Ash in coke can be guaranteed within the required range only if the ash in mixed coal is controlled, and the ash in coke is generally 1.3-1.4 times that in mixed coal. So the ash of mixed coal in general is controlled between 9% and 10% [1].

**2.1.3. Volatiles ( $V_{daf}$ ).** The metamorphism of coal can be reflected through volatiles. The volatiles can be calculated roughly on a weighted average basis according to the volatile of single coal. The product yield of coking gas and chemical products can be improved by appropriately adding high volatile coal, with the best content being controlled in the range of 24%–30% [14].

**2.1.4. Sulfur Content ( $S_{t,d}$ ).** Sulfur content is the harmful component in coke, most of which is brought into the blast furnace by coke. The sulfur content of pig iron is directly affected by it, and the quality of pig iron is declined. Therefore, the control of sulfur content is indispensable. Related study [15] shows that 60%-70% sulfur content from coal is transferred into coke, the inorganic sulfur in coal is transferred into coke as sulfocompounds, and the remaining part stays in ash in the form of sulfate and sulfide. The coke rate of mixed coal is 70%-80% usually, while the sulfur content in coke is 80%-90% of that in coal [16]. Thus, the sulfur content of the mixed coal should be constrained in the range of 0.6%-0.7%.

**2.1.5. Caking Property.** Caking property refers to the ability of coal forming plastic substance during coking, which serves as a necessary condition in coking process and affects the coking property of mixed coal [17]. It commonly can be reflected by three indexes, X, Y, and G in glial layer. Among them, shrinkage X can be used to estimate whether the final shrinkage of coal can cause difficulty of coke pushing, the maximum thickness of the glial layer Y indicates the amount of liquid produced with the mixed coal, and the bond index G, to a certain extent, reflects the content of the gum. So, the caking property of mixed coal can be measured by the mentioned indicators.

**2.2. Index Assessing the Coke Quality.** Coke provides heat and reducing gas, occupying an important position in the furnace. The hot strength (coke reactivity index CRI and coke strength after reaction CSR) is a main index in quality judgment of coke, and so are ash, sulfur, and volatile [15, 17].

**2.2.1. Ash ( $A_d$ ).** Ash in coke is mainly composed of  $\text{SiO}_2$ ,  $\text{Al}_2\text{O}_3$ , and other acidic oxide composition with high melting point. In the blast furnace refining slag discharge requires a lot of solvent to reduce the melting point of the compound [18]. The ash content of the coke all comes from mixed coal. The higher the ash content, the less the carbon content of the coke, the more solvent required, the more slag, and the lower yield [19]. The ash in coke is an inert component, and it will reduce the caking property of mixed coal, increase the coke crack, and reduce the coke mechanical strength. At the same time, the alkali metal oxide in the ash component also catalyzes the  $\text{CO}_2$  reaction of the coke and increases the  $\text{CO}_2$  reaction rate.

**2.2.2. Sulfur Content ( $S_{t,d}$ ).** 80%–85% of sulfur content in coke, as one of indicators affecting coke quality, comes from the mixed coal. It influences the sulfur content in steel from the blast furnace smelting and the design of the blast furnace and destroys the environment severely [20].

**2.2.3. Hot Strength of Coke.** The coke supplies heat and carbon in the blast furnace, serving as a reducing agent for the furnace reaction; its hot strength is a mechanical strength

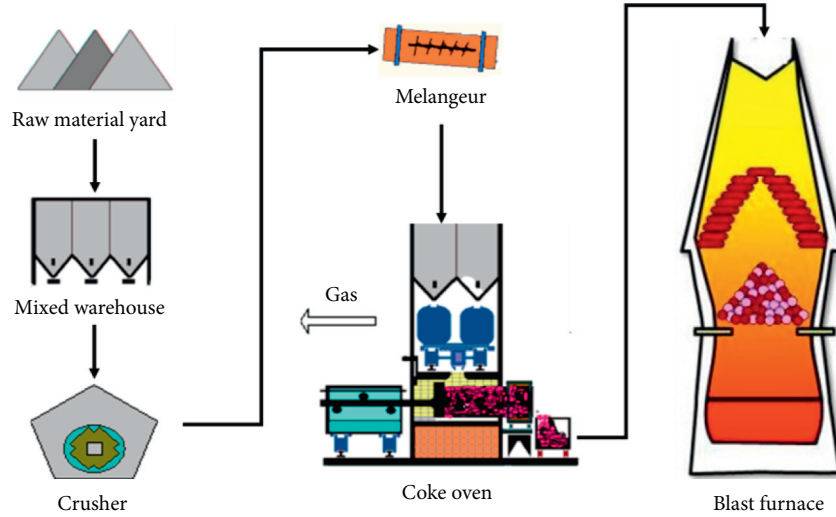


FIGURE 1: Coke production.

TABLE 1: The range of content of mixed coal corresponding indicators and the impact on coke production.

| Index of mixed coal | Range (%)    | Effect  |
|---------------------|--------------|---|
| $M_t$               | $0 << 8$     | Bad for processing and transportation, deteriorates thermal stability and thermal conductivity when burning, and reduces the coke yield and extends the coking cycle when coking. |
| $A_d$               | $9 << 10$    | Leads to sudden drop of the cold state strength, the increase of residue in a blast furnace, and finally production reduction.  |
| $V_{daf}$           | $24 << 30$   | The product yield of coking gas and chemical products can be improved by appropriately adding high volatile coal.   |
| $S_{t,d}$           | $0.6 << 0.7$ | The sulfur content of pig iron is directly affected by it, and the quality of pig iron is declined.   |

index reflecting the thermal performance [21]. This indicator characterizes the ability to resist crushing and abrasion subjected to thermal stress and mechanical force at specific temperature and atmosphere. The thermal performance is commonly denoted by coke reactivity index (CRI) and coke strength after reaction (CSR). The lower the CSI and the higher the CSR, the better the thermal performance of coke. At the same time, they are the most critical parameters used to evaluate the high temperature performance of coke [22].

### 3. Neural Networks Based on Domain Knowledge

Neural networks are often merely trained and generalized in data-driven mechanisms. However, it helps to speed up the search of approximation and improve the prediction quality if adding priori knowledge (including knowledge such as symmetry, invariance common sense, and so on) of the desired approximation function to networks. Domain knowledge focuses on the investigation of important issues or concepts in the located fields, as well as the interrelationships of them, which can make up for the shortcomings of unclear directivity and unexplainable results.

Based on the current situation that high-quality coking coal resources are relatively scarce, a multicoal source coking method is proposed. Considering that the nonlinear

relationship between parameters of mixed coal and coke determined by the inherent properties of coking environment, process and chemical reaction, cannot be described accurately by functions simply, neural networks are very suitable for coke quality prediction in coking systems, which avoids the function description between system characteristic indexes and dependent variables, gets relationship between input and output according to memory and feature extraction, and has features of generalization, distributed knowledge storage, associative memory, and parallel processing.

#### 3.1. Multilayer Feedforward Backpropagation (FB) Network.

The pretreated samples are randomly presented to the neural network in traditional algorithm, and the output signals of neurons are revealed through forward recursion next [23]:

$$y_q = f(v_q) = f(u_q - \theta_q) = \left( \sum_{j=1}^n w_{qj} x_j - \theta_q \right). \quad (1)$$

The parameters are corrected in the negative gradient direction in terms of the selected error energy function  $\zeta(n) = f(e_j(n))$  after obtaining the error signal  $e_j(n) = d_j(n) - y_j(n)$  from comparing the expected value and the actual output [23].

$$\begin{aligned}
w_{ji}^{(s)}(k+1) &= w_{ji}^{(s)}(k) + \eta^{(s)} \delta_j^{(s)} y_j^{(s-1)}, \\
\delta_j^{(s)} &= \begin{cases} e_j(n) \phi_j'(v_j(n)), & j \text{ is the putput node,} \\ \phi_j'(v_j(n)) \sum_k \delta_k(n) w_{kj}(n), & j \text{ is the hidden node.} \end{cases} \quad (2)
\end{aligned}$$

Another algorithm is investigated owing to the slow convergence of the original one, which can be roughly divided into the following:

- (1) Heuristic improved algorithms, including back-propagation with momentum updates, search convergence method, and bulk update of variable learning rates
- (2) Numerical optimization techniques, including conjugate gradient backpropagation, recursive back-propagation based on least squares, and backpropagation with adaptive activation function

The concepts of sensitive area and its width of Sigmoid function  $f()$  are discussed in literature [24]; each implicit neuron has a corresponding respective sensitive area; directions of neurons depend on that of the hidden node. The hidden nodes interact with each other in the feature space formed by themselves, which affects the training performance of neural network. The input interval of hidden node whose output is within  $(0, a)$  is defined as the  $a$ -level sensitive area  $A_a$  of the node [25]:

$$A_a = \{x \in R^{n \times 1} : 0 < f(x^T w + \theta) < a\}, \quad (3)$$

where  $x$  is the vector inputted into synapse, and its weights are  $x_j, j = 1, 2, \dots, n, x = [x_1, x_2, \dots, x_n]^T$ .  $x_j$  is connected to the neuron  $q$  through the weight of synapses  $w_{qj}$ ; the nonlinear activation function  $f()$  is able to regulate the amplitude of outputs and enhance abilities of classification, function approximation, anti-noise jamming, and so on;  $\theta$  is the threshold to reduce the cumulative input of the activation function.

The width of  $A_a$  is defined as [25]

$$G_a = P_a - P_0 = \frac{\ln(a/(1-a))}{\|w\|}, \quad (4)$$

where  $P_a$  and  $P_0$  represent the equivalent hyper-surface when outputs  $f(x^T w + \theta)$  of hidden nodes are  $a$  and  $0$ , respectively, and  $\|w\|$  is the  $L_2$  norm of matrix  $w$  (also known as Euclidean norm).

**3.2. Radial Basis Function (RBF) Network.** The function approximation of the multilayer perceptron is realized by nesting the weighted sums, and the RBF is itself a general approximator following the interpolation theory, and it is delicate and tight in the mathematical point of view. Different from the random approximation of multilayer perceptions, RBF based on radial basis functions includes two stages in solving nonlinear mapping problems: high-dimensional transformation and least squares estimation of input samples. The idea is to map the samples non-linearly to

high-dimensional space where the single weighted sum is performed and eventually outputs results. Its rationality is explained by the configurable Cover theorem. The classical training method is to train the hidden layer by using the K-means clustering algorithm in unsupervised mode and then to calculate the weight vector of the output layer using the recursive least squares method. The method has the characteristics of simple calculation and accelerated convergence. The following calculation is carried out once the samples are given:

- (1) It is assumed that the training set  $\{\varphi(i), d(i)\}_{i=1}^N$  and the number of iterations  $n = 1, 2, \dots, N$  are known already. Initialize  $\hat{w}(0) = 0, p(0) = \lambda^{-1}I$ , where  $\lambda$  is a small positive number.
- (2) For a given encoder  $C$ , minimize the total clustering variance  $\sum_{j=i}^K \sum_{C(i)=j} \|x_i - \hat{\mu}_j\|^2$  according to the clustering mean  $\{\hat{\mu}_j\}_{j=1}^K$ , resulting in the clustering mean  $\{\hat{\mu}_j\}_{j=1}^K$ , and minimize the encoder  $C(i) = \arg \min_{1 \leq j \leq K} \|x(i) - \hat{\mu}_j\|^2$ . Repeat the step until the clustering result remains unchanged.
- (3) Calculate  $\mathbf{p}(n) = \mathbf{p}(n-1) - (\mathbf{p}(n-1)\Phi(n)\Phi^T(n) \mathbf{p}(n-1) / 1 + \Phi^T(n)\mathbf{p}(n-1)\Phi(n))$ , where  $p(n) = R^{-1}(n)$ , and  $R(n)$  is the  $K \times K$  order related function of the hidden units:  $\mathbf{R}(n) = \sum_{i=1}^n \phi(x_i)\phi^T(x_i)$ ,  $\phi(x_i) = [\varphi(x_i, x_1), \varphi(x_i, x_2), \dots, \varphi(x_i, x_k)]^T$ ; among them,  $\varphi(x_i, x_j) = \exp(-(1/2\sigma_j^2)\|x_i - x_j\|^2)$   $j = 1, 2, \dots, K$ .
- (4)  $g(n) = p(n) \cdot \Phi(n)$ .
- (5)  $a(n) = d(n) - \hat{\mathbf{w}}^T(n-1)\Phi(n)$ .
- (6)  $\hat{\mathbf{w}}(n) = \hat{\mathbf{w}}(n-1) + \mathbf{g}(n)a(n)$ : the weight vector of the output layer is obtained.

**3.3. Algorithm to Better the Neural Network Generalization.** The training process of the network can be regarded as the process of constructing the fitting curve. The network is of pretty good generalization when the mapping calculated by it is correct although inputs are beyond the training samples. There appears overfitting phenomenon with redundant network structure in which the extra synapses memorize unexpected characteristics of information. It can be readily seen that the wider the sensitive area defined by (4) is, the stronger the node generalization ability is. In addition, the ideal function is the smoothest function of approximation and mapping functions for a given error, taking up fewer computing resources. And the process of seeking for it is called network pruning. In this section, a regularized network training algorithm based on the sensitive area of Sigmoid function is developed, and the specific process is shown in Figure 2.



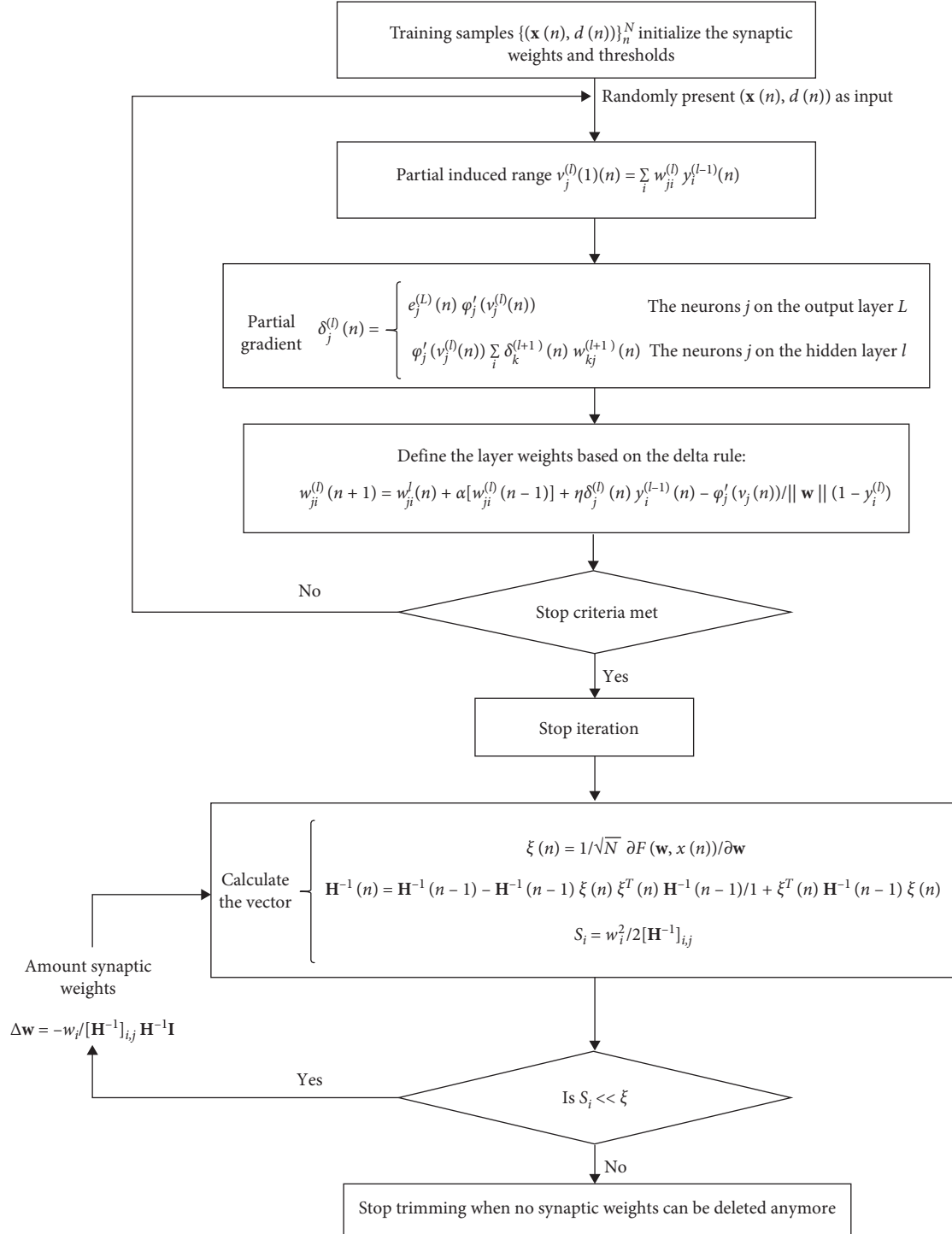


FIGURE 2: The regularized network training algorithm based on the sensitive area of Sigmoid function.

The sensitive area of the activation function proposed in equation (4) is introduced to the mean square error performance function:

$$\varsigma(n) = \frac{1}{2} \sum_{k \in C} e_k^2(n) - G_a(n), \quad (5)$$

where the neuron  $k$  is an output node.

The wider the sensitive area is, the stronger the generalization ability of the node is. The parameters in the parallel network are modified considering the distribution of sensitive areas and the anti-jamming ability of the network. Then excess weights, that is, the weights having little impact on performance and leading to overtraining, are removed in the pruning process using the second-order information of the error function and considering the complexity and the

training error performance. The final parameters then make it possible to minimize the growth of performance functions when removing them away, which are ideal compromises considering complexity and error performance, and they can further enhance generalization capabilities.

The training is actually a fitting process of nonlinear inputs and outputs, and the generalization of network can be seen as the nonlinear interpolation of the verification data. It is said that the network loses the ability to generalize in other samples when overfitting occurs. Numerous hidden units in a cascade network are likely to store too much noise effect. In that case, it is of necessity to use cross-validation.

Check the verification error in the corresponding parameters state under the set viewing cycle, and enter the next cycle if passing the validation. The training error converges along with the increase of training times, while the validation error monotonically decreases and then rises. The network starts to capture the noise information after crossing the minimum point. Hence, the state is regarded as the stopping criterion to reduce the occurrence of overfitting.

## 4. Coke Quality Prediction Model

**4.1. Data Preprocessing.** In general, the direct input of untreated data is not optimal. For example, the finite limit of the logistic activation function is  $(0, 1)$ , but the sample value is tremendous compared with its finite limit, which results in the function being almost saturated and the training being stagnated. In addition, the backpropagation algorithm is similar to the LMS algorithm, of which the calculation time is heavily dependent on the condition number  $\lambda_{\max}/\lambda_{\min}$ . The  $\lambda_{\max}/\lambda_{\min}$  of nonzero mean input is bigger than that of the zero-mean input.

Hence, it is required that mean value of the data on the entire training set is close to 0. The conventional processing is the mean centering and variance regulation. Let the input and output modes  $A \in R^{n \times m}$ ,  $C \in R^{p \times m}$  be arranged by columns, calculate the mean value of the  $n$ -th row of  $A$  and the  $p$ -th row of  $C$ , and subtract this mean value in each row. Calculate the variance of the  $n$ -th row in  $A$  and the  $p$ -th row in  $C$ , and divide the corresponding values in each row. The processing of the input matrix  $A$  and the output pattern matrix  $C$  should be synchronous.

**4.2. Multilayer Feedforward Backpropagation Network.** The parameters of mixed coal are selected as inputs and the coke quality as outputs in the FB model shown in Figure 3; cascade forward network can be used to solve more complex problems and improve the training precision, in which each subsequent layer is connected to the input layer and the adjacent one and the output is directly influenced by the input layer.

The excitation function is used to process the result of the summer, and the nonlinear function is generally used to maximize the efficiency of the network. In this paper, the logsig and tansig functions are used to perform the combination trial and error.

In the cascade forward backpropagation (CF) structure, it can be seen that the accuracy and fit effect that the training method can obtain are better than that of other training functions; hence, it is fixed as the training function in this paper.

The number of hidden nodes depends on the number of training samples, the size of the noise, and rules hidden in data. Commonly, it is best for the number of hidden nodes to be twice the number of input layers. It is illustrated by lots of tests carried out for 12–16 nodes in a single hidden layer that 16 nodes produce the best prediction result.

**4.3. RBF.**  $N$  samples are given in advance and the parameter  $r$  specifies the division between the training set  $\ell'$  and the verification one  $\ell''$ . Then, there are  $(1-r)N$  samples in  $\ell'$  and  $rN$  samples in  $\ell''$ . It is a good choice that the parameter  $r$  remains at a fixed value of 0.2 according to the simulation experiment by Keams; that is, the training set is assigned 80% of the samples and the remaining samples are distributed in the validation set  $\ell''$ .

The number of input nodes depends on the dimension of the input vector  $x$ , written as  $m_0$ .

The hidden layer size  $m_1$  is determined by the number of planned clusters, controlling the network performance and the computational complexity.

The clustering mean  $\hat{\mu}_j$  is obtained via the clustering algorithm and serves as the center of the basis function  $\varphi(x_j)$  which is known as  $x_j$ .

$\sigma = d_{\max}/2K$  is the extended parameter, where  $d_{\max}$  denotes the maximum distance between the centers ensuring that the hidden layer unit is neither too sharp nor too flat.

## 5. Simulation Results and Analysis

800 sets of mixed coal quality parameters and their corresponding coke quality indexes are randomly selected from the 1000 sets of preprocessed data to train FB, CF, and RBF networks, respectively.

**5.1. Results and Analysis of Training Results.** RBF can meet the conditions required by the Cover theorem. The adjustment of weight and threshold is considered by functions newrbe and newrb in constructing the network, so there are no specific training and learning functions for the network. In that case, only the training results of FB and CF networks are compared through error curves in Figures 4 and 5:

As can be seen from the figure, the results of training, verification, and test of the two networks basically remain consistent. Under almost identical iteration number, the performance of cascade forward network reaches 0.01, showing a slightly better behavior than that of the forward network (0.05) with the training time being close, while the prediction errors of both networks are comparatively small on the whole.

**5.2. Results and Analysis of Prediction Results.** The remaining 200 sets of mixed coal quality parameters are

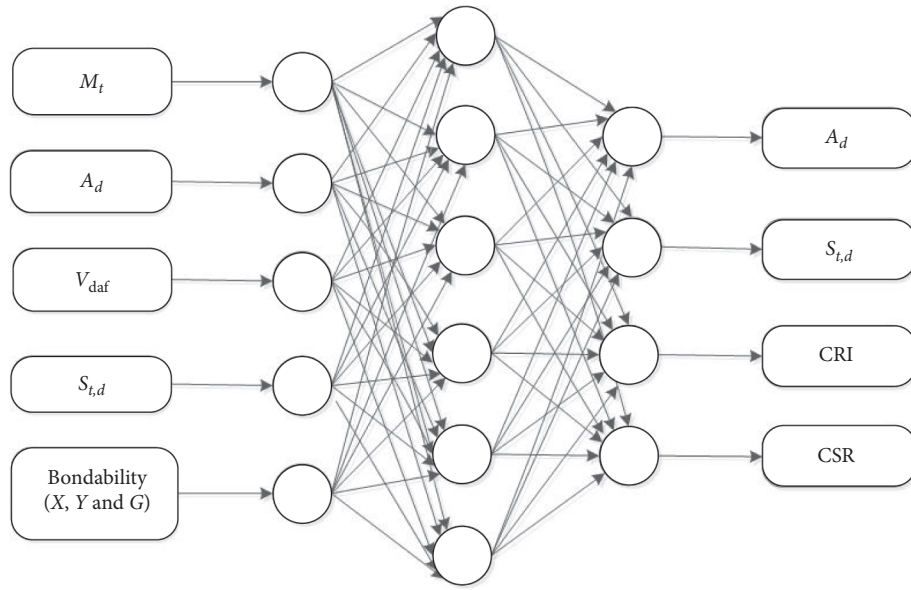


FIGURE 3: The structure of FB network.

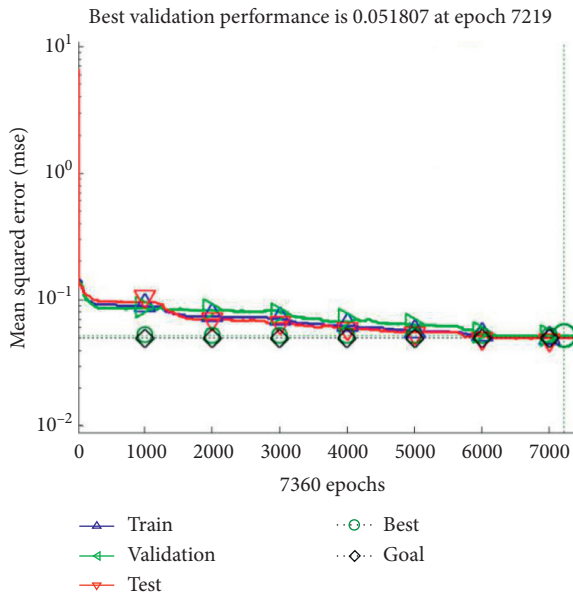


FIGURE 4: The training error curve of FB network.

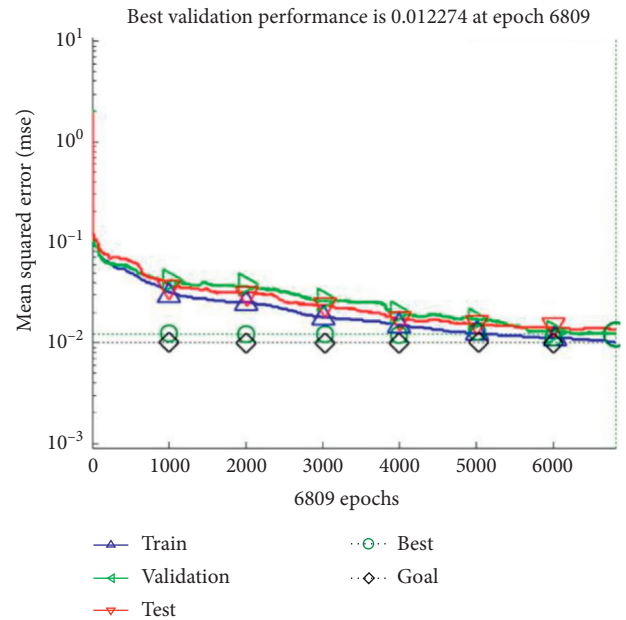


FIGURE 5: The training error curve of CF network.

utilized to predict the coke quality by the above trained networks, and predictions of ash, sulfur, CRI, and CSR are shown in Figures 5–8. Prediction errors of 5% and 10% represent different levels of accuracy required by the business.

The line with the symbol of “◁” in Figures 6–9 denotes the error margin of 10%, The line with the symbol of “▷” in Figures 6–9 denotes the error margin of –10%, the line with the symbol of “△” represents that of 5%, the line with the symbol of “▽” represents that of -5% and the line with the symbol of “○” represents that of 0%. The symbols of “×”, “-,” and “|” show the prediction results of the FB, CF, and RBF networks, respectively.

Only three sets of ash prediction error are more than 10% in the FB, CF, and RBF networks, while the error between 5% and 10% exists in 51, 29, and 48 sets, respectively. There are 13 and 12 sets in the FB and the RBF networks, respectively, while there is one set in the CF network with the prediction error being more than 8%. It can also be illustrated from Figure 6 that the predictive values of CF network are closer to the 0% error margin line as a whole compared to the results of FB and RBF networks. Hence, the conclusion obviously can be drawn that the CF network performs much better than the others in terms of ash prediction.

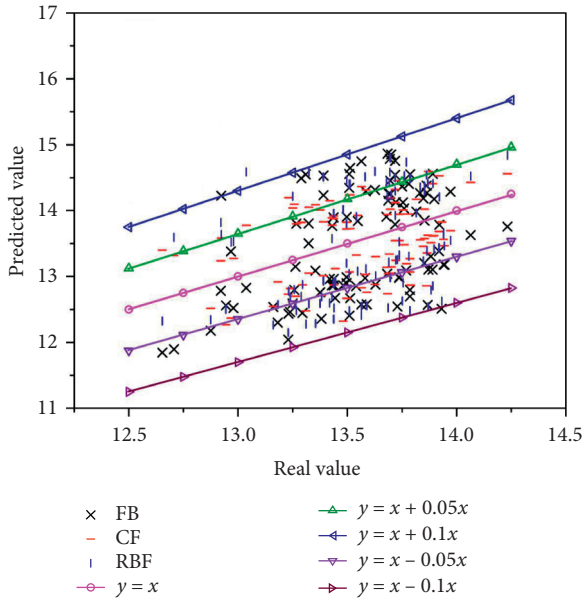
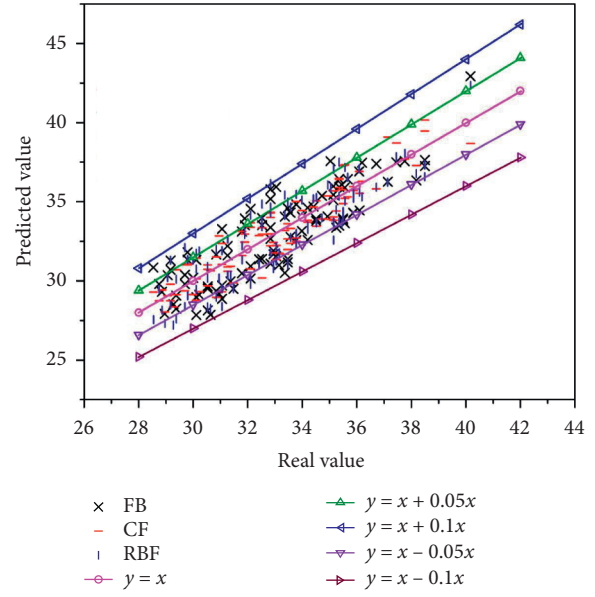
FIGURE 6: Comparison of prediction errors of  $A_d$ .

FIGURE 8: Comparison of prediction errors of CRI.

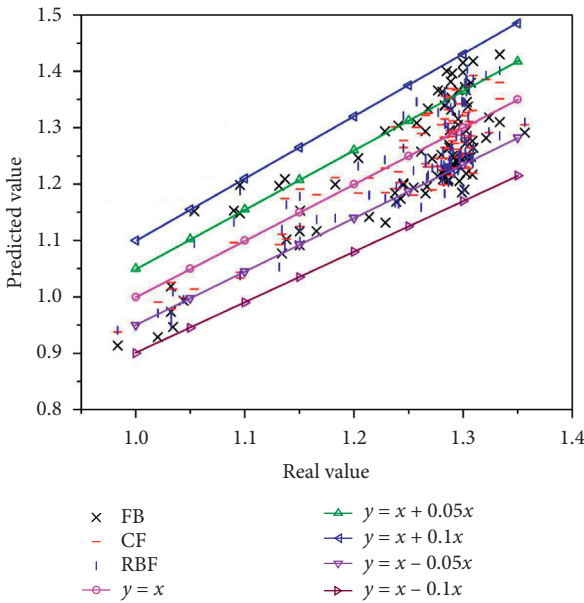
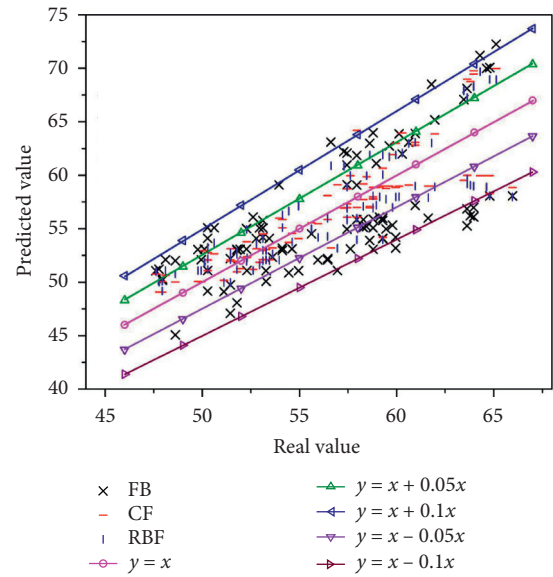
FIGURE 7: Comparison of prediction errors of  $S_{r,d}$ .

FIGURE 9: Comparison of prediction errors of CSR.

The prediction errors of sulfur content of three networks are all within 10%, and errors of 54, 17, and 27 sets are beyond 5%, respectively. The forecast of sulfur content is more excellent than that of ash and the RBF is better than BP for prediction. It can also be seen from Figure 7 that the number of the predictive values of CF network outside the 5% error margin line is significantly less than that of FB and RBF networks. The CF network behaves excellently in the completion of sulfur content prediction.

The errors are also controlled within 10% in the three networks as for the CRI index, 29, 10, and 32 sets of samples among them having the errors over 5%. The prediction effect of CRI is much better than that of the former indicators. It is shown that the number of the predictive values of FB and

RBF networks outside the 5% error margin line is significantly greater than that of CF network and a majority of errors of CF are within the 5% error margin line.

The prediction of FB network is not so good for CSR, with 10 sets having errors between 10% and 15%, while there are only 2 sets in two other networks with the same errors. With 62, 19, and 24 sets of errors over 5%, respectively, in the three networks, the overall prediction effect of CRI is worst among the predictions of the four properties. It is shown that most of the predictive values of FB network are distributed between the 5% and 10% error margin line, and the effect is comparatively poor; values of CF network near the yellow line are slightly more than that of RBF. Hence, the CF network is quite suitable for the prediction of the data structure selected in this paper.

It can be concluded in general that the prediction results for the four properties are satisfactory and the prediction accuracy is relatively high, with the best prediction effect for the CRI and the most disappointing predictive result for the CSR. The cascade forward network performs best among the three networks with most errors controlled within 5%, while the forward BP behaves the worst, as its most prediction errors vary from 5% to 10% and even over 10%.

## 6. Conclusions

Three networks for coke quality prediction were proposed based on the forward backpropagation network, the cascade forward backpropagation network, and the radial basis network and the corresponding results were compared, respectively. It is shown that the prediction of the CRI is closest to the actual values with errors within 10% in all three networks, better than that of the CSR with errors between 10% and 15% in 14 sets of samples. The prediction errors for the four properties are all within 15% and it can be said that the three networks all have a relatively high prediction accuracy. When it comes to the network structure, the cascade forward network behaves best with errors mostly controlled within 5%, which acts as an excellent guide during the coke production to some extent. It is the additional connection between each layer and the front ones that ensures the prediction accuracy. The closer the hidden layer is to the output layer, the greater the amount of information of effect on weights adjusting. And the traditional forward network is not suitable here as its most errors are between 5% and 10% and even over 10%. In this way, it is concluded that the cascade forward network matches the data structure selected in this paper.

However, the main budget constraint is computing time for large-scale samples, and the advantage of cascade networks with a large number of internal connections is likely to become no longer obvious; hence, large-scale learning problems need to be further investigated.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Authors' Contributions

All authors contributed equally to this work.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 51974131), Science and Technology Project of Hebei Education Department (no. BJ2017021), NCST Natural Science Funds for Distinguished Young Scholars (no. JQ201711), Hebei Province Natural Science Fund for Excellent Young Scholars (no.

E2018209248), Hebei Provincial Postgraduate Demonstration Course Project in 2020 (no. KCJSX2020053), and NCST Project Establishment and Construction of Postgraduate Demonstration Course.

## References

- [1] H. P. Tiwari, S. K. Haldar, A. Roy, A. Bhattacharjee, and P. Mishra, "Data mining—new perspectives on predicting coke quality in recovery stamp charged coke making process," *Metallurgical Research & Technology*, vol. 112, no. 6, p. 603, 2015.
- [2] X. Tang, S. Snowden, B. C. McLellan, and M. Höök, "Clean coal use in China: challenges and policy implications," *Energy Policy*, vol. 87, pp. 517–523, 2015.
- [3] H. V. Jagadish, J. Gehrke, A. Labrinidis et al., "Big data and its technical challenges," *Communications of the ACM*, vol. 57, no. 7, pp. 86–94, 2014.
- [4] X. Jin, B. W. Wah, X. Cheng, and Y. Wang, "Significance and challenges of big data research," *Big Data Research*, vol. 2, no. 2, pp. 59–64, 2015.
- [5] S. C. Chelgani, S. S. Matin, and J. C. Hower, "Explaining relationships between coke quality index and coal properties by Random forest method," *Fuel*, vol. 182, pp. 754–760, 2016.
- [6] R. Roest, H. Lomas, K. Hockings, and M. R. Mahoney, "Fractographic approach to metallurgical coke failure analysis. Part 1: cokes of single coal origin," *Fuel*, vol. 180, pp. 785–793, 2016.
- [7] K. Shahzad, S. Kanwal, S. Nawaz, N. Sheikh, and S. Munir, "Effects of moisture and coal blending on the hardgrove grindability index of Pakistani coals," *International Journal of Coal Preparation and Utilization*, vol. 34, no. 1, pp. 1–9, 2014.
- [8] J. P. Mathews, V. Krishnamoorthy, E. Louw et al., "A review of the correlations of coal properties with elemental composition," *Fuel Processing Technology*, vol. 121, pp. 104–113, 2014.
- [9] L. Zhang, W. Liu, and D. Men, "Preparation and coking properties of coal maceral concentrates," *International Journal of Mining Science and Technology*, vol. 24, no. 1, pp. 93–98, 2014.
- [10] R. Morga, I. Jelonek, K. Kruszezwska, and W. Szulik, "Relationships between quality of coals, resulting cokes, and micro-Raman spectral characteristics of these cokes," *International Journal of Coal Geology*, vol. 144–145, pp. 130–137, 2015.
- [11] B. D. Flores, A. G. Borrego, M. A. Diez et al., "How coke optical texture became a relevant tool for understanding coal blending and coke quality," *Fuel Processing Technology*, vol. 164, pp. 13–23, 2017.
- [12] M. A. Diez, R. Alvarez, and C. Barriocanal, "Coal for metallurgical coke production: predictions of coke quality and future requirements for coke making," *International Journal of Coal Geology*, vol. 50, no. 1–4, pp. 389–412, 2002.
- [13] C. Liu, Z. Xie, F. Sun, and L. Chen, "Exergy analysis and optimization of coking process," *Energy*, vol. 139, pp. 694–705, 2017.
- [14] J. Chen, G. Liu, H. Li, and B. Wu, "Mineralogical and geochemical responses of coal to igneous intrusion in the Pansan coal mine of the Huainan coalfield, Anhui, China," *International Journal of Coal Geology*, vol. 124, pp. 11–35, 2014.
- [15] S.-F. Zhang, L.-Y. Wen, K. Wang, C. Zou, and J. Xu, "Effects of additives on sulfur transformation, crystallite structure and properties of coke during coking of high-sulfur coal," *Journal of Iron and Steel Research International*, vol. 22, no. 10, pp. 897–904, 2015.

- [16] R. M. Garbarino and R. T. Tonti, "Desulfurization and its effect on calcined coke properties," in *Essential Readings in Light Metals*, pp. 119–122, Springer International Publishing, Cham, Switzerland, 2016.
- [17] S. S. Raaj, S. Arumugam, M. Muthukrishnan, S. Krishnamoorthy, and N. Anantharaman, "Characterization of coal blends for effective utilization in thermal power plants," *Applied Thermal Engineering*, vol. 102, pp. 9–16, 2016.
- [18] R. Loison, P. Foch, and A. Boyer, *Coke: Quality and Production*, Elsevier, Amsterdam, Netherlands, 2014.
- [19] A. Ahmadpour, N. Jahanshahi, S. Rashidi, N. Chenarani, and M. J. D. Mahboub, "Application of artificial neural networks and adaptive neuro-fuzzy inference systems to predict activated carbon properties for methane storage," *Adsorption Science & Technology*, vol. 32, no. 4, pp. 275–290, 2014.
- [20] L. C. Edwards, K. J. Neyrey, and L. P. Lossius, "A review of coke and anode desulfurization. Essential readings in light metals," *Electrode Technology for Aluminum Production*, vol. Volume 4, pp. 130–135, 2013.
- [21] S. S. Matin and S. C. Chelgani, "Estimation of coal gross calorific value based on various analyses by random forest method," *Fuel*, vol. 177, pp. 274–278, 2016.
- [22] S. S. Haykin, *Neural Networks and Learning Machines*, Pearson, Upper Saddle River, NJ, USA, 2009.
- [23] M. R. Patil and P. Sawant, "Corporate social responsibility, performance and sustainability reporting of shree cement company in India: a case study," *Journal of High Technology Management Research*, vol. 3, no. 4, pp. 74–87, 2014.
- [24] E. Hadavandi, J. C. Hower, and S. C. Chelgani, "Modeling of gross calorific value based on coal properties by support vector regression method," *Modeling Earth Systems and Environment*, vol. 3, no. 1, p. 37, 2017.
- [25] L. Ljung, "System identification," in *Signal Analysis and Prediction*, pp. 163–173, Birkhäuser Boston, Basel, Switzerland, 1998.

## Research Article

# A Research on the Control System of High-Speed Homopolar Motor with Solid Rotor Based on Flywheel Energy Storage

Lili Jing,<sup>1</sup> Yandong Yu ,<sup>2</sup> and Xiaochuan Xue<sup>3</sup>

<sup>1</sup>Key Laboratory of High Speed Signal Processing and Internet of Things Technology Application, Jining Normal University, Ulanqab, Inner Mongolia 012000, China

<sup>2</sup>Department of Computer Science, Jining Normal University, Ulanqab, Inner Mongolia 012000, China

<sup>3</sup>Systems Engineering Research Institute of China State Shipbuilding Corporation, Beijing 100191, China

Correspondence should be addressed to Yandong Yu; [cfssyyd@163.com](mailto:cfssyyd@163.com)

Received 17 April 2020; Revised 13 June 2020; Accepted 17 June 2020; Published 17 July 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Lili Jing et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In view of the defects of the motors used for flywheel energy storage such as great iron loss in rotation, poor rotor strength, and robustness, a new type of motor called electrically excited homopolar motor is adopted in this paper for flywheel energy storage. Compared to general motors, this motor has the advantages of simple structure, high rotor strength, and low iron loss in rotation. A double closed-loop PI governing system of the new motor was designed, modeled, and simulated with this motor as the controlled object on simulation platform. The simulation result shows that the PI-controlled electrically excited homopolar motor system realized fast speed governing and good system performance, achieved the purpose of quick charge by flywheel energy storage, verified the correctness and effectiveness of the design scheme, and provided data reference and theoretical basis for the practice of future flywheel energy storage experiments.

## 1. Introduction

Today the whole world is confronted with an increasing shortage of energy. Besides, constantly developing new energies, how to utilize existing energies effectively has become a hot spot of people's attention. Energy storage is an important way to achieve efficient utilization of energy [1], so the storage technology of electric energy, as a main form of energy consumed today, has become the most important research subject in the energy field.

In the 21st century, flywheel energy storage, owing to its many advantages including high energy ratio, high power ratio, high efficiency, wide application, no pollution, no noise, easy maintenance, and long service life, is bound to bring about an evolution to energy storage and show the prospect of green energy storage technology [2]. In overseas, the research of flywheel energy storage technology started between 1950s and 1970s when magnetically levitated energy

storage flywheel was created in the form of a series of prototype. At present, international leading companies in the field of flywheel energy storage include US-based Active Power which produces and supplies with UPS and energy storage flywheel developed by Japanese company with HTS magnetic levitation technology, which is mainly used for the trolley bus. As a replacement of traditional energy technology, flywheel energy storage technology is applied to fields such as transportation, energy generation, and aerospace relying on its strong competitiveness. Now, China is relatively backward in this field and still in the fledging period, but is gradually emerging being impacted by the current great upsurge around the world. The Flywheel Energy Storage Laboratory of Tsinghua University was the first institution in China to initiate the research of this technology and developed an energy storage flywheel prototype in 1999, which laid a solid theoretical and experimental basis for our flywheel energy storage technology. In

addition, Beijing University of Aeronautics and Astronautics, Nanjing University of Aeronautics and Astronautics and North China Electric Power University, etc., also completed this technology experiment and obtained their unique patented products.

Flywheel energy storage is a kind of energy-storing device utilizing physical interconversion between electric energy and mechanical energy (kinetic energy). It is mainly composed of auxiliary devices including flywheel, flywheel motor, motor controller, bearings, and vacuum chamber [3]. The flywheel energy storage system completes energy absorption and release by the flywheel motor and its controller, so the flywheel motor is the main part of this device. Systems using different flywheel motors and control types may have different performances. The energy storage flywheel motor and its control system is a complex integrated electromechanical system. As a new type of motor, the homopolar motor has not been widely known yet. Due to its advantages such as simple structure, high-strength rotor, and low iron loss in high-speed rotation, it is especially suitable for energy storage flywheel. By modeling and simulation research of the high-speed control system of electrically-excited homopolar motor [4] as the controlled object, we aim to provide a preliminary theoretical foundation and essential experimental basis for the Research on Topology Optimization and Dynamically Decoupling Nonlinear Control of High-speed Permanent Magnet Homopolar Motor with Solid Rotor by NSF of Inner Mongolia, in order to save the design cost and R&D time.

## 2. Structure and Principle of Homopolar Motor

Electrically-excited homopolar motor, such as general motors, also consists mainly of rotor, stator, and other relevant components, as shown in Figure 1. The differences are the stator and rotor materials and structural design. Due to cost consideration, the rotor bearings choose angular contact ball bearings in order to ensure stable and high-speed operation. Its rotor is made of 40CrNiMoA, adopts high-strength forged solid-steel structure with flywheel-motor integrated design, and is designed with uniformly distributed concaves and convexes [5] and antisymmetric upper and lower ends, as shown in Figure 2. The flywheel-motor integrated design, which enables the rotor to integrate the functions of kinetic energy storage and electromagnetic force output, greatly simplifies the design process, reduces the production cost, and makes the solid-steel rotor; the flywheel has the same level of energy storage density as the composite rotor flywheel. Its stator is composed of two parts, the armature and the exciter; its section view is shown in Figure 3. The motor stator has such characteristics: its armature windings use tinned copper braid which is parallel to rotor length axis and evenly distributed along rotor circumference; currents in adjacent in-phase windings flow to opposite directions; and three-phase current generates space rotating magnetic field; while the salient pole of the rotor induced by electromagnetic force drives the motor to rotate. The motor's parameters are provided in Table 1.

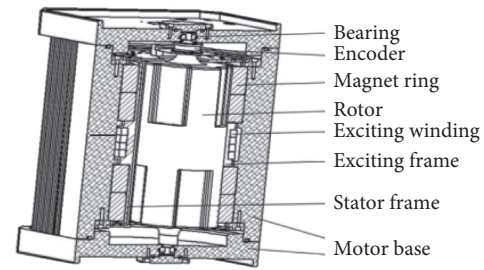


FIGURE 1: Section of motor assembly.

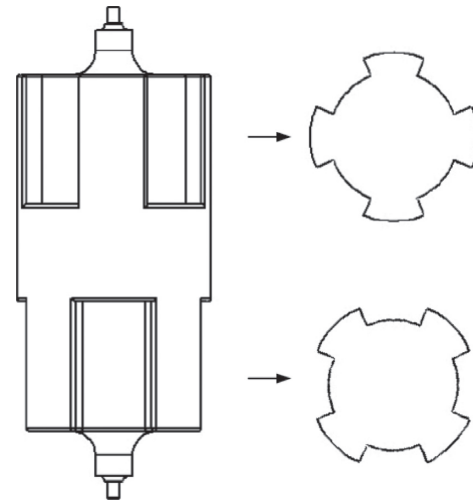


FIGURE 2: Structure of solid rotor.

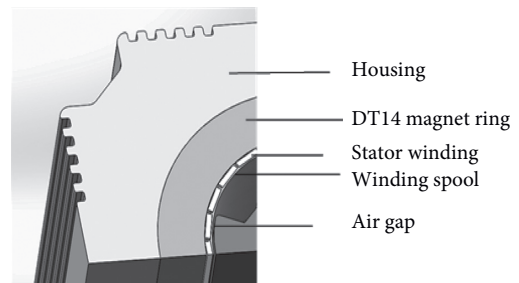


FIGURE 3: Section of stator.

Running principle of electrically excited homopolar motor; its space magnetic field is generated by the impressed current through the exciting windings around the end circumference. Its magnetic loop is a closed circuit formed by exciting flux flowing from the upper end of the rotor, through the magnet ring and air gap into the lower end. Because the rotor's magnetism is generated by the exciting coil, its magnetic strength is controllable by exciting current. The motor rotates as its rotor's salient pole senses the corresponding space rotating magnetic field, while the required space rotating magnetic field is generated by the variable three-phase current through the rotor. The motor's operational principle is shown in Figures 4 and 5. As stator current flows from A to C, seen from the motor's upper end in Figure 4, A-phase winding current flows



TABLE 1: Parameters of electrically excited homopolar motor.

| Motor parameters | Rotor parameters                |                | Stator parameters |                   | Exciting parameters |                       |                         |
|------------------|---------------------------------|----------------|-------------------|-------------------|---------------------|-----------------------|-------------------------|
| Housing material | 6063-t5                         | Rotor material | 40CrNiMoA         | Armature material | VF wire             | Magnet ring material  | DT4-E                   |
| Magnetic gap     | 4.0 mm                          | Diameter       | 90 mm             | Winding diameter  | 1.82 mm             | Magnetic conductivity | 0.0113H·m <sup>-1</sup> |
| Housing volume   | 200 × 200 × 230 mm <sup>3</sup> | Length         | 213 mm            | Stator turns      | 10                  | Coil turns            | 1250                    |
| Gross mass       | 19 kg                           | Rotor mass     | 5.98 kg           | Rated current     | 10.4 A              | Coil current          | <2 A                    |

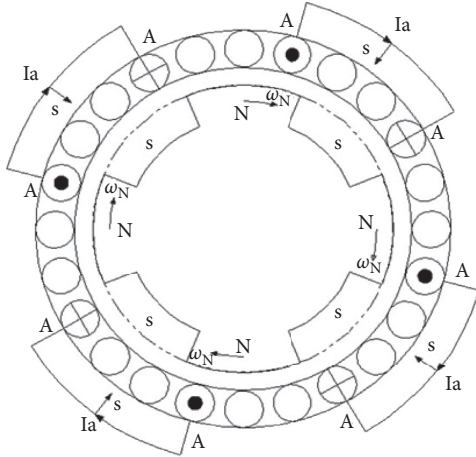


FIGURE 4: Upper stator and rotor of the motor.

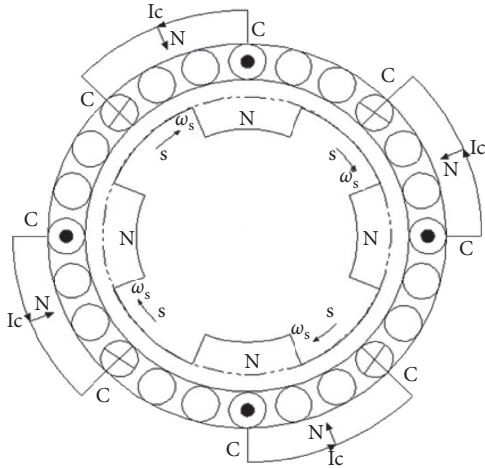


FIGURE 5: Lower stator and rotor of the motor.

clockwise; the magnetic field generated by armature winding is inward at S pole and outward at N pole. The upper salient pole (N pole) of the rotor rotates clockwise when induced by this electromagnetic force; similarly and simultaneously, seen from the motor's lower end in Figure 5, the rotor's salient pole (S pole) also rotates clockwise being driven by the magnetic field generated by the current varying counterclockwise in C-phase winding, i.e., the rotating magnetic field produced by variable current in the motor stator acts on the motor rotor and drives the latter to rotate.

### 3. Mathematical Model of Homopolar Motor

As the rotor of the motor in this paper has the characteristics of the salient pole structure, we built and deduced a unified dynamic mathematical model [6, 7] of the motor in the stator's three-phase coordinate system based on the basic electromagnetic relationship. And because the counter electromotive force wave and phase current output by the motor are similar to those by brushless DC motor, this motor's mathematical model is approximate to that of brushless DC motor.

The mathematical model of electrically excited homopolar motor using the motor's own vector in the stator's three-phase coordinate system is shown in Figure 6.

In this figure  $as$ ,  $bs$ , and  $cs$  are the corresponding axes of three-phase stator windings;  $A-A'$ ,  $B-B'$ , and  $C-C'$  are three-phase stator windings; the rotor speed  $\omega_r$  is set as counterclockwise,  $\psi e$  is the flux linkage vector generated by the external electromagnetic, forming an angle of  $\alpha$  with the plane of the stator three-phase coordinate system; and  $\psi e_0$  is the projection of  $\psi e$  in the coordinate system plane, with an angle of  $\theta$  with axis  $as$ .

Without affecting the control performance, simplified analysis shows that three-phase windings of the motor are symmetric, and its stator's voltage state equation is as follows:

$$\begin{bmatrix} u_a \\ u_b \\ u_c \end{bmatrix} = \begin{bmatrix} R_a & 0 & 0 \\ 0 & R_b & 0 \\ 0 & 0 & R_c \end{bmatrix} \begin{bmatrix} i_a \\ i_b \\ i_c \end{bmatrix} + \begin{bmatrix} L_a & L_{ab} & L_{ac} \\ L_{ba} & L_b & L_{bc} \\ L_{ca} & L_{cb} & L_c \end{bmatrix} \frac{d}{dt} \begin{bmatrix} i_a \\ i_b \\ i_c \end{bmatrix} + \begin{bmatrix} e_a \\ e_b \\ e_c \end{bmatrix}. \quad (1)$$

In this equation,  $u_a$ ,  $u_b$ , and  $u_c$  are, respectively, the stator's three-phase terminal voltages (V);  $R_a$ ,  $R_b$ , and  $R_c$  are the resistances ( $\Omega$ ) of the motor's three-phase windings;  $L_{ab}$ ,  $L_{ba}$ ,  $L_{ca}$ ,  $L_{ac}$ ,  $L_{bc}$ , and  $L_{cb}$  are the mutual inductances (H) among the three-phase stator windings;  $L_a$ ,  $L_b$ , and  $L_c$  are the self-inductances (H) of the motor's three-phase windings;  $e_a$ ,  $e_b$ , and  $e_c$  are, respectively, the counterelectromotive forces (V) of the motor's three-phase windings; and  $i_a$ ,  $i_b$ , and  $i_c$  are the three-phase currents (A) of the stator.

Assuming that the three-phase winding is symmetric and the reluctance of rotor does not change as the rotor position shifts.

Namely,  $R = R_a = R_b = R_c$ ,  $M = L_{ab} = L_{ac} = L_{bc} = L_{ba} = L_{ca} = L_{cb}$ , and  $L = L_a = L_b = L_c$ .

In the brushless DC motor, the three-phase winding is symmetric; then,

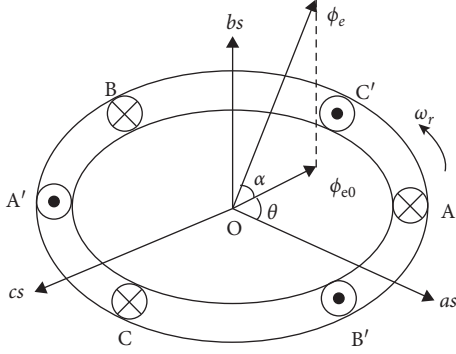


FIGURE 6: Mathematical model of motor's 3-phase stator.

$$i_a + i_b + i_c = 0, \quad (2)$$

$$Mi_a + Mi_b + Mi_c = 0. \quad (3)$$

Thus, from equation (1), we can deduce

$$\begin{bmatrix} u_a \\ u_b \\ u_c \end{bmatrix} = \begin{bmatrix} R & 0 & 0 \\ 0 & R & 0 \\ 0 & 0 & R \end{bmatrix} \begin{bmatrix} i_a \\ i_b \\ i_c \end{bmatrix} + \begin{bmatrix} L-M & 0 & 0 \\ 0 & L-M & 0 \\ 0 & 0 & L-M \end{bmatrix} \frac{d}{dt} \begin{bmatrix} i_a \\ i_b \\ i_c \end{bmatrix} + \begin{bmatrix} e_a \\ e_b \\ e_c \end{bmatrix}. \quad (4)$$

The flux-linkage equation of the homopolar motor stator in the three-phase coordinate system is as follows:

$$\begin{cases} \psi_A = L_A i_A + M_{AB} i_B + M_{AC} i_C + \psi_e \cos \alpha \cos \theta, \\ \psi_B = L_B i_B + M_{AB} i_A + M_{BC} i_C + \psi_e \cos \alpha \cos \left( \theta + \frac{2}{3} \pi \right), \\ \psi_C = L_C i_C + M_{AC} i_C + M_{BC} i_B + \psi_e \cos \alpha \cos \left( \theta + \frac{4}{3} \pi \right). \end{cases} \quad (5)$$

The equation of the electromagnetic torque generated by the interaction between impressed exciting field and stator current and the motor's motion equation is as follows:

$$T_e = \frac{(e_a i_a + e_b i_b + e_c i_c)}{\omega_r}, \quad (6)$$

$$J \frac{d\omega_r}{dt} = T_e - T_L - B\omega_r. \quad (7)$$

In the equations,  $T_e$  is the motor's electromagnetic torque ( $N \cdot m$ );  $\omega_r$  is the motor's mechanical angular velocity (rad/s);  $J$  is the motor's rotational inertia ( $\text{kg} \cdot \text{m}^2$ );  $T_L$  is the load torque ( $N \cdot m$ ); and  $B$  is the damping coefficient ( $\text{s}^{-1}$ ).

## 4. Control Strategy of Homopolar Motor

Famous scholar Astrom once said [8] "in process control today, more than 95% of the control loops are of PID type." It shows the dominance [9] of the PID controller in the field of process control. The flywheel energy storage motor control system focuses more on the motor's speed regulation time and less on indexes such as control precision. Therefore, PID strategy is used to control the double closed-loop system of the homopolar solid rotor motor and applied in the research on flywheel energy storage.

**4.1. PID Control Strategy.** PID control is the most commonly used control law in the simulation control system. It is a kind of linear controller. Its structure is shown in Figure 7.  $r(t)$  is set input value;  $e(t)$  is control deviation;  $u(t)$  is control variable; and  $y(t)$  is actual output value.

PID control law:

$$u(t) = K_p e(t) + \frac{K_p}{T_i} \int_0^t e(t) dt + K_p T_d \frac{de(t)}{dt}, \quad (8)$$

or

$$u(t) = K_p e(t) + K_i \int_0^t e(t) dt + K_d \frac{de(t)}{dt}. \quad (9)$$

In the form of transfer function,

$$G_s = \frac{U(s)}{E(s)} = K_p + \frac{K_p}{T_i s} + K_p T_d s, \quad (10)$$

or

$$G_s = \frac{U(s)}{E(s)} = K_p + \frac{K_i}{s} + K_d s. \quad (11)$$

In the equations above,  $K_p$  is proportionality coefficient;  $T_i$  is integral time constant;  $T_d$  is differential time constant;  $K_i$  is integral coefficient,  $K_i = (K_p/T_i)$ ; and  $K_d$  is differential coefficient,  $K_d = K_p \times T_d$ .

For the convenience of computer implementation, the discrete method is adopted to convert continuous PID to the corresponding digital PID control algorithm [10, 11]. The expression is as follows:

$$u(k) = K_p e(k) + \frac{K_p T}{T_i} \sum_{j=0}^k e(j) + \frac{K_p T_d}{T} [e(k) - e(k-1)], \quad (12)$$

or

$$u(k) = K_p e(k) + K_i T \sum_{j=0}^k e(j) + \frac{K_d}{T} [e(k) - e(k-1)]. \quad (13)$$

In the equations,  $T$  is sampling period;  $k$  is sampling signal, ( $k = 1, 2, 3, \dots$ );  $e(k)$  is system deviation input value at the  $k$ th time of sampling;  $e(k-1)$  is system deviation input value at the  $(k-1)$ th time of sampling; and  $u(k)$  is control quantity output value at the  $k$ th time of sampling.

Its pulse transfer function is

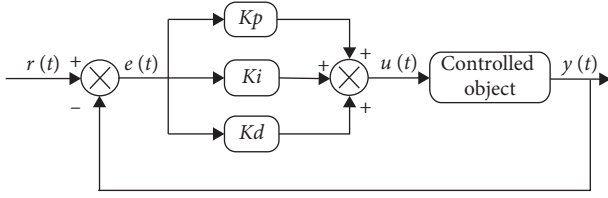


FIGURE 7: Block diagram of the PID controller.

$$G(z) = \frac{U(z)}{E(z)} = K_p + \frac{K_i T}{1 - z^{-1}} + \frac{K_d}{T} (1 - z^{-1}). \quad (14)$$

Commonly used digital PID control algorithms include the following:

- (1) Full PID control: this kind of control algorithm directly presents the execution point of the actuator (such as motor speed and valve opening), and thus it is named full variable or positional PID control algorithm. It directly calculates based on the defining equations (9) and (10) of PID control law. Its disadvantage is that the output is full variable and always associated with all past states; moreover, it requires accumulation of  $e(k)$  in the calculation, which increases a huge amount of work.
- (2) Incremental PID control [11]: this control algorithm applies where the output is only the increment of the control variable. Where the actuator needs the increment of the controlled variable (such as step motor), incremental PID control algorithm can be deduced from full variable PID control algorithm. The formula is

$$\Delta U(k) = Ae(k) + Be(k-1) + Ce(k-2). \quad (15)$$

In this equation,  $A = K_p + K_i T + (K_d/T)$ ;  $B = K_p - (2K_d/T)$ ; and  $C = (K_d/T)$

Once the sampling period and coefficients  $A$ ,  $B$ , and  $C$  are determined, we can get the increment of the controlled variable by just entering three times of measurement deviations before and after the measurement. It presents the positional variation of the actuator between the two times of sampling. Compared with the positional one, the incremental one requires a much smaller amount of calculation and simple algorithm and is convenient for software programming. Therefore, it is widely used in practical production. Based on the incremental PID control algorithm, the recursion formula of full quantity PID control algorithm can also be deduced:

$$U(k) = U(k-1) + \Delta U(k). \quad (16)$$

The above formula is the digital recursive PID control algorithm widely used in digital control systems today.

Another great advantage of digital PID control is that it allows us to improve PID depending on the object and working condition so that PID control can achieve high-quality control effect in different fields. Commonly used improved digital PID control algorithms [12–15] include

differential forward PID control algorithm, incomplete differential control algorithm, integral clamping PID control algorithm, PID control algorithm with dead zone, and integral separation PID control algorithm.

**4.2. PID Controller Tuning.** We can see from PID control principle diagram 6 that the PID controller is composed of three major control links (proportion, integral, and differential) decided mainly by PID's three parameters which are mutually inseparable, interconnected, and restrictive. Look at the roles of proportion, integral, and differential in terms of overshoot, stability, response, and steady-state precision of the control system [10, 11]:

- (1) Proportional control link: proportionality coefficient  $K_p$  and deviation signal  $e(t)$  together constitute the proportional control link. It is the most basic link and plays a leading role in PID control. Once deviation occurs to the system, this link is immediately activated to reduce deviation. However, proportional control alone cannot eliminate system deviation. There will be a residual error. Larger proportional control coefficient  $K_p$  means stronger control effect, faster system response, and higher stability precision, but a too large  $K_p$  value is easy to cause system overshoot, instability, and oscillation. On the contrary, a too small  $K_p$  will reduce system's stability precision, prolong the transition time for system to reach the set value, slow down its response time, and spoil system static and dynamic characteristics.
- (2) Integral control link: integral coefficient  $K_i$  and the time integral of deviation together constitute the integral control link. Its main function is to eliminate system static error and improve system stability. Larger  $K_i$  means faster speed of eliminating system static error, but a too large  $K_i$  value is easy to lead to overshoot at the beginning of control due to integral saturation. On the contrary, a too small  $K_i$  value will have little effect in eliminating system static error and will also reduce the system's control precision.
- (3) Differential control link: differential coefficient  $K_d$  and the time differential of deviation together constitute the differential control link. It compensates the function of the proportional control link, prevents system overshoot, suppresses oscillation, and maintains system stability. It mainly predicts deviation change in the process of response and reduces it. The  $K_d$  value should not be too large; otherwise, it will suppress the response process in advance, prolong the tuning time, and reduce the system's antiinterference performance.

From the above effects of proportional coefficient  $K_p$ , integral coefficient  $K_i$ , and differential coefficient  $K_d$  on system and the synergy of the three, we can see that they are all critical for system control. The parameters of the PID controller vary with the steady-state condition of the control system. Once the working condition (or the process object's "characteristic") changes, then the optimal value of the

controller parameter changes too. This means controller tuning may be necessary at any time so that the control system and process control object reach the required rapidity, precision, and stability [16]. Commonly used PID parameter tuning methods include the following.

Testing method parameter tuning is also called engineering tuning method. It applies the valuable experience people accumulated in long-term engineering practice to the control system to test directly. This method is simple and easy to grasp. The testing method is to find out characteristic parameters in the response system according to the testing result and work out the controlling parameter with empirical formula.

Attenuation curve methods include 4:1 attenuation curve method and 10:1 attenuation curve method.

Empirical method, also known as trial and error, is a PID parameter tuning method including three steps from proportion to integral and then to differential. Turn the parameters repeatedly according to this sequence until a desired effect is achieved. Here is the procedure. First, tune the value of the proportionality coefficient. Increase it one by one and find one with small overshoot and quick response. Now, observe whether there is a static error. Add the integral coefficient if there is. When tuning the integral coefficient, decrease the set value from large to small and at the same time tune the proportionality coefficient. Observe the system response until the system static error is eliminated. While the eliminating system static error, keep an eye on system overshoot and response time. If the system overshoot is very large or response is slow, add the differential coefficient and tune again until desired control effect is obtained.

## 5. Control System Design and Modeling

Flywheel energy-storage system is composed of three major parts, flywheel rotor, motor/generator, and controller. The controller, as the system command center, directly determines the system's control performance [17]. Therefore, in the process of designing the control system of high-speed homopolar solid rotor motor with energy-storing flywheel, we built a control system simulation model (Figure 8) of electrically excited homopolar motor based on its operation principle and mathematical model with the modular modeling method [18–20] on MATLAB/SIMLINK simulation platform in order to shorten the design cycle and lower the research cost and risk. The control system simulation model of homopolar motor consists of seven subsystems, including homopolar motor, inverter, speed controller PI-W1, current controller PI-current, PWM generator, commutator, and decoder, as well as speed reference  $w_r$ -ref and load module TL.

As energy-storing flywheel requires high motor efficiency, the system of this homopolar motor is designed with double-closed-loop control. The outer one is speed loop. It applies PI control strategy to its speed controller PI-W1 to ensure the system dynamic tracking performance and suppress speed fluctuation. The inner one is the current loop. It applies integral separation PI control strategy to its current controller PI-current to limit the maximum current, makes

the system have enough accelerating torque, and ensures system stability and high-performance operation. Signal from the current controller is sent into the PWM generator which then produces PWM signal with adjustable pulse width. Then, PWM signal, along with rotor position decoder signal, is sent into commutator module which generates Hpwm-Lon mode of pulse modulation waves, according to rotor position decoding (Table 2) and commutation rule (Table 3) to control ordered switching of the inverter's IGBT full bridge circuits. This driving mode is an output mode with high leg PWM modulation and low leg constant live. It can suppress motor torque ripple and improve system dynamic performance [21]. Details of each subsystem are shown in Figures 9–13.

From the six switching states of the homopolar motor upon commutation, we note that the motor's commutation sequence is associated with the rotor position. According to the motor's operational principle, when the motor rotates clockwise, the switching relationship between the rotor position and the switching tube is given in Table 2. The switching tube's conducting state is set to "1", while its breaking state is set to "0".

## 6. Simulation Result and Analysis

Acceleration energy storage is one of the core issues of system control and relates to system's energy storing speed and safety. When energy storing acceleration is controlled, the flywheel motor runs as an electric motor [17]. In this paper, the parameters required by the homopolar motor and controller were set in a homopolar motor control system model which was built on the "simulation model editor" interface on software platform Matlab, see Table 4.

In order to test each performance index of the system after all module parameter settings of the homopolar motor control system and the corresponding program file embeddings were completed, we repeated system running, simulation, and tuning again and again and finally obtained the simulation result and made an necessary analysis of the tuning result and waveform.

In the simulation experiment of tuning the homopolar motor control system, in order to ensure the motor's safe startup and simulate some practical factors in its real environment, the motor started with a load (0.1 N·m) at the initial preset speed of  $n^* = 15000$  r/min. At 0.5 S it added an external load  $TL = 0.4$  N·m and continued to operate. The speed response wave is as shown in Figure 14.

Figure 14 shows that the motor quickly accelerates to 15,000 r/min after the start-up, with a little speed overshoot at the beginning, yet  $\sigma\% < 0.04\%$ , acceleration time  $t_p = 0.137$  S, and transition time  $t_s = 0.23$  s, with very small and negligible static error. At 0.5 S, the motor begins to run with load ( $TL = 0.4$  N·m) and slightly slows but maintains basically at 15,000 r/min. The whole system responds fast and runs steadily. Eventually, the motor speed is basically stable at the set value with little fluctuation.

Other performance indexes of the motor: Figure 15 is the torque wave of the electric motor. At 0.137 S, the torque is basically stable at 2.1 N·m but with notable fluctuation.

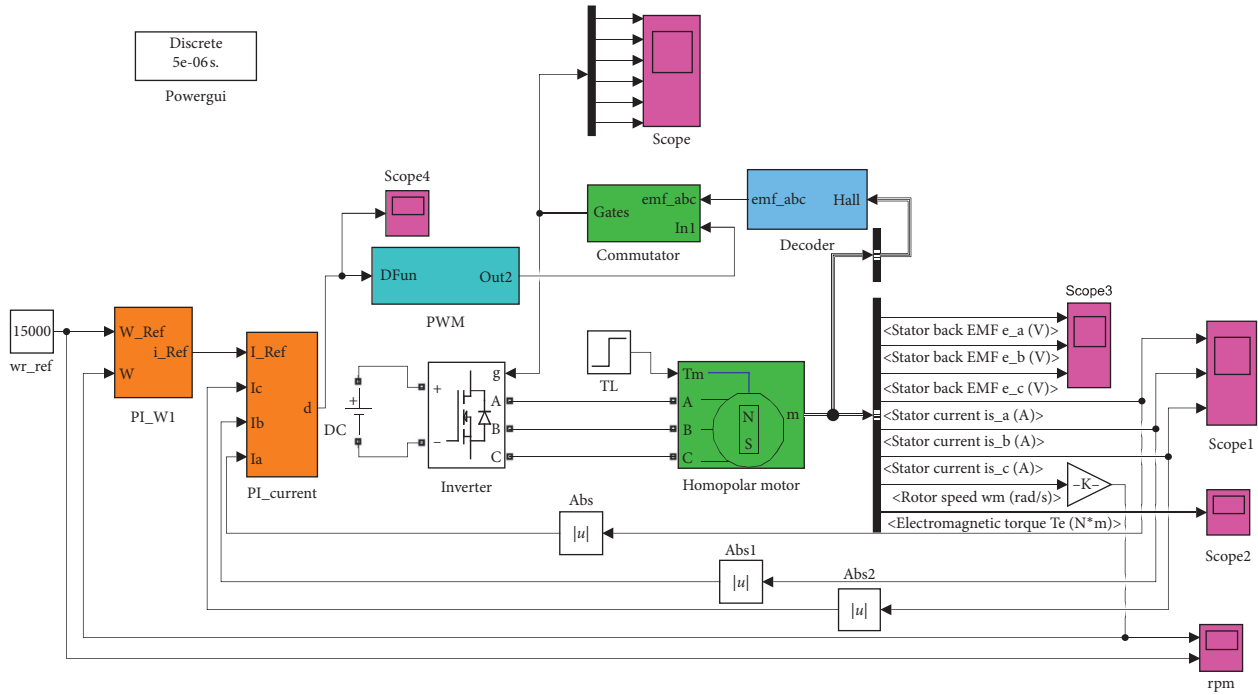


FIGURE 8: Simulation model of the homopolar motor control system.

TABLE 2: HALL signal decoding of rotor position.

| <i>ha</i> | <i>hb</i> | <i>hc</i> | <i>emf-a</i> | <i>emf-b</i> | <i>emf-c</i> |
|-----------|-----------|-----------|--------------|--------------|--------------|
| 0         | 0         | 0         | 0            | 0            | 0            |
| 0         | 0         | 1         | 0            | -1           | +1           |
| 0         | 1         | 0         | -1           | +1           | 0            |
| 0         | 1         | 1         | -1           | 0            | +1           |
| 1         | 0         | 0         | +1           | 0            | -1           |
| 1         | 0         | 1         | +1           | -1           | 0            |
| 1         | 1         | 0         | 0            | +1           | -1           |
| 1         | 1         | 1         | 0            | 0            | 0            |

TABLE 3: Rotor position and commutation rule of switching tube conducting and breaking.

| <i>emf-a</i> | <i>emf-b</i> | <i>emf-c</i> | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 |
|--------------|--------------|--------------|----|----|----|----|----|----|
| 0            | 0            | 0            | 0  | 0  | 0  | 0  | 0  | 0  |
| 0            | -1           | +1           | 0  | 0  | 0  | 1  | 1  | 0  |
| -1           | +1           | 0            | 0  | 1  | 1  | 0  | 0  | 0  |
| -1           | 0            | +1           | 0  | 1  | 0  | 0  | 1  | 0  |
| +1           | 0            | -1           | 1  | 0  | 0  | 0  | 0  | 1  |
| +1           | -1           | 0            | 1  | 0  | 0  | 1  | 0  | 0  |
| 0            | +1           | -1           | 0  | 0  | 1  | 0  | 0  | 1  |
| 0            | 0            | 0            | 0  | 0  | 0  | 0  | 0  | 0  |

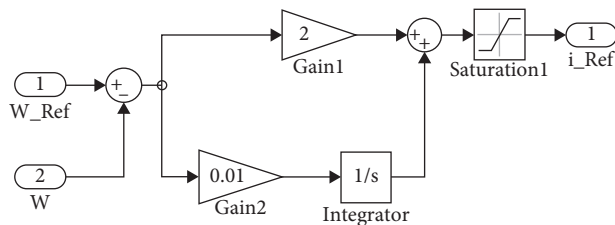


FIGURE 9: Subsystem simulation model of speed controller PI-W1.

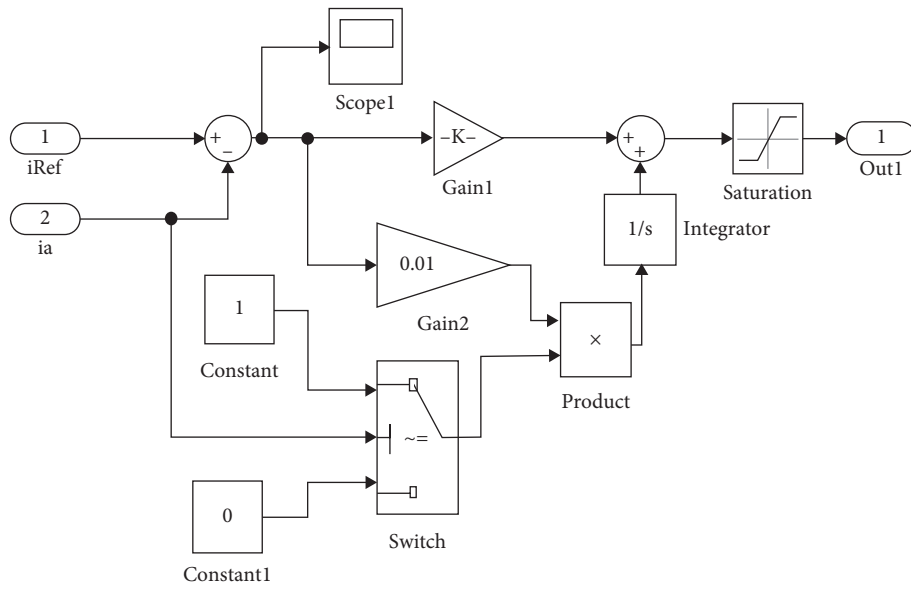


FIGURE 10: Subsystem simulation model of current controller PI-current.

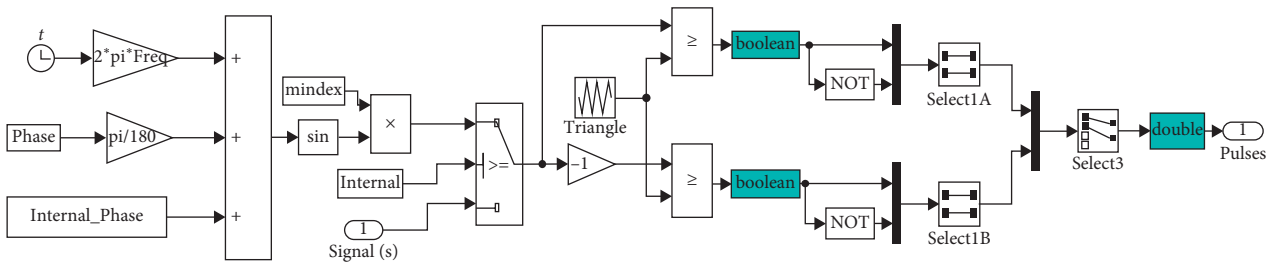


FIGURE 11: Subsystem simulation model of the PWM generator.

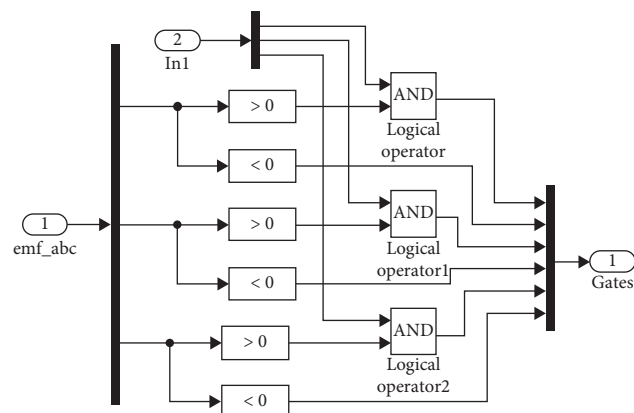


FIGURE 12: Subsystem simulation model of the commutator.

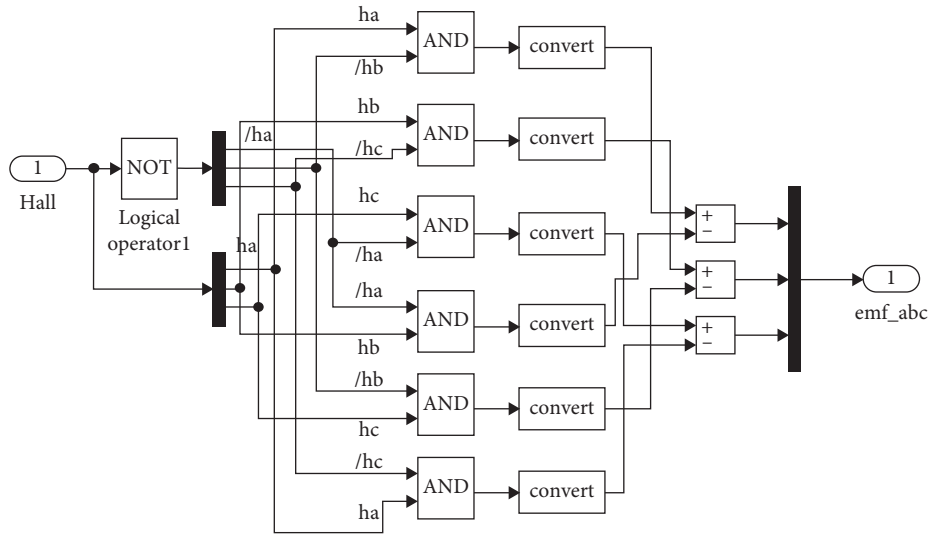


FIGURE 13: Subsystem simulation model of the decoder.

TABLE 4: Parameter list of the homopolar motor control system.

|                       |  |                                       |
|-----------------------|--|---------------------------------------|
|                       | Stator resistance $R_s = 0.055 \Omega$                         | Constant torque $0.047 \text{ N.m/A}$ |
|                       | Stator inductance $L_s = 0.000115 \text{ H}$                   | Static friction $0.4 \text{ N.m}$     |
| Homopolar motor       | Moment of inertia<br>$J = 5.615 \times 10^{-3} \text{ kg.m}^2$ | Number of pole pairs $p = 4$          |
|                       | System viscous damping $10^{-3} \text{ N.ms}$                  | Rated speed $8000 \text{ rpm}$        |
| PI speed controller   | $K_p = 2$  | $K_i = 0.01$                          |
| Current PI controller | $K_p = 0.25$   | $K_i = 0.01$                          |

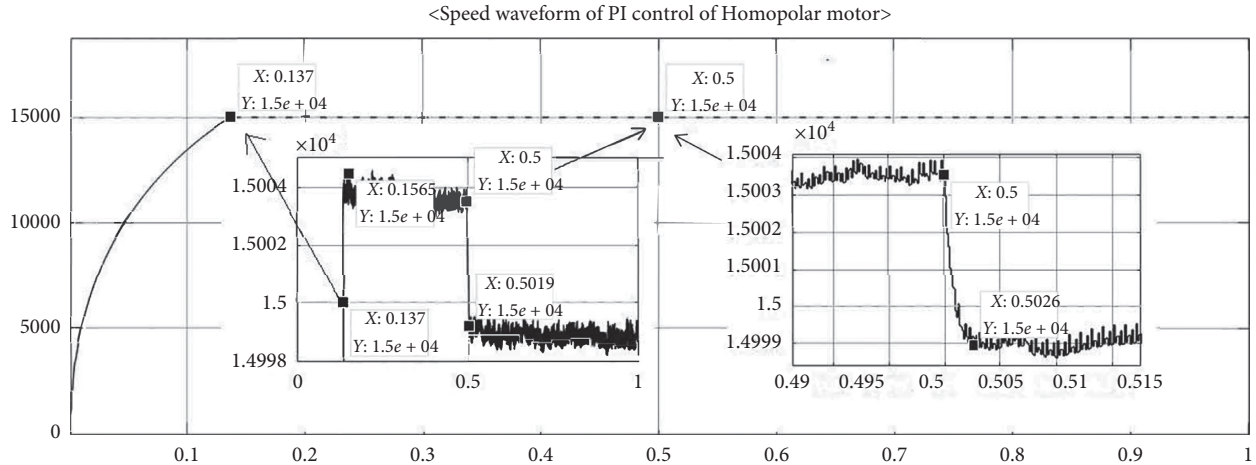


FIGURE 14: Speed waveform of PI control of homopolar motor.

When the motor runs with load at 0.5 S, the torque changes to 2.5 N.m. The pulse wave output by the commutator is PWM signal in Hpwm-Lon modulation mode (see Figure 16). Figure 17 is the induction electromotive force of three-phase motor stator. The stator's induction electromotive force is in the form of 180° trapezoidal wave and each corresponding phase of the magnetic flux is also in

trapezoidal waveform, so this system is also known as the speed-governing system of trapezoidal-wave permanent magnet synchronous motor (TPMSM). Figure 18 is the induced current waveform of the three-phase motor stator. The motor current is in 120° rectangular waveform. As the voltage adopts PWM chopper control, the rectangular wave contains a significant harmonic component. Upon the

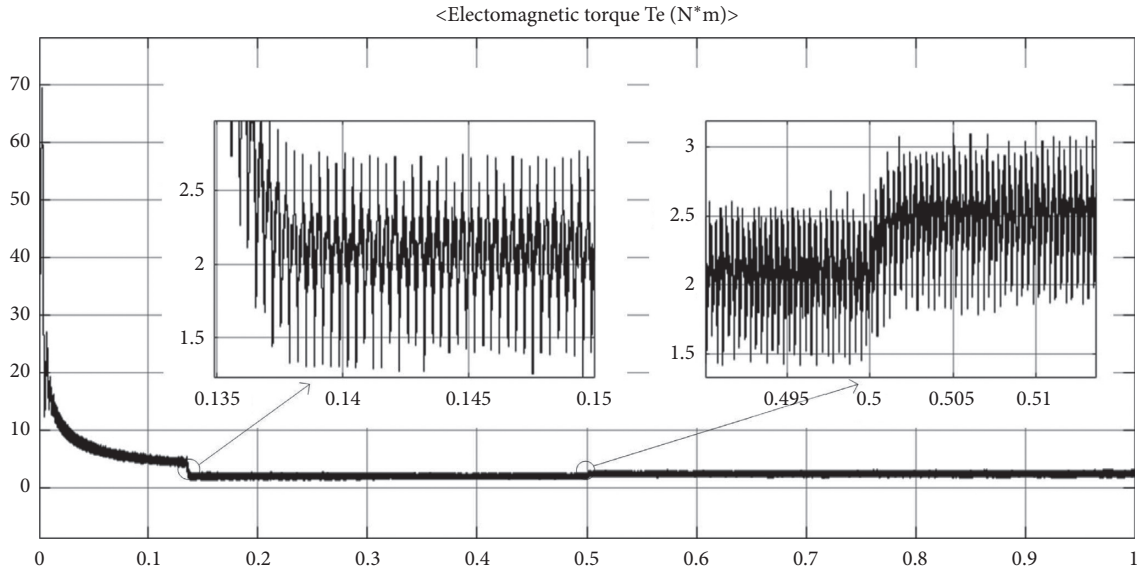


FIGURE 15: Torque waveform of the homopolar motor of PI control.

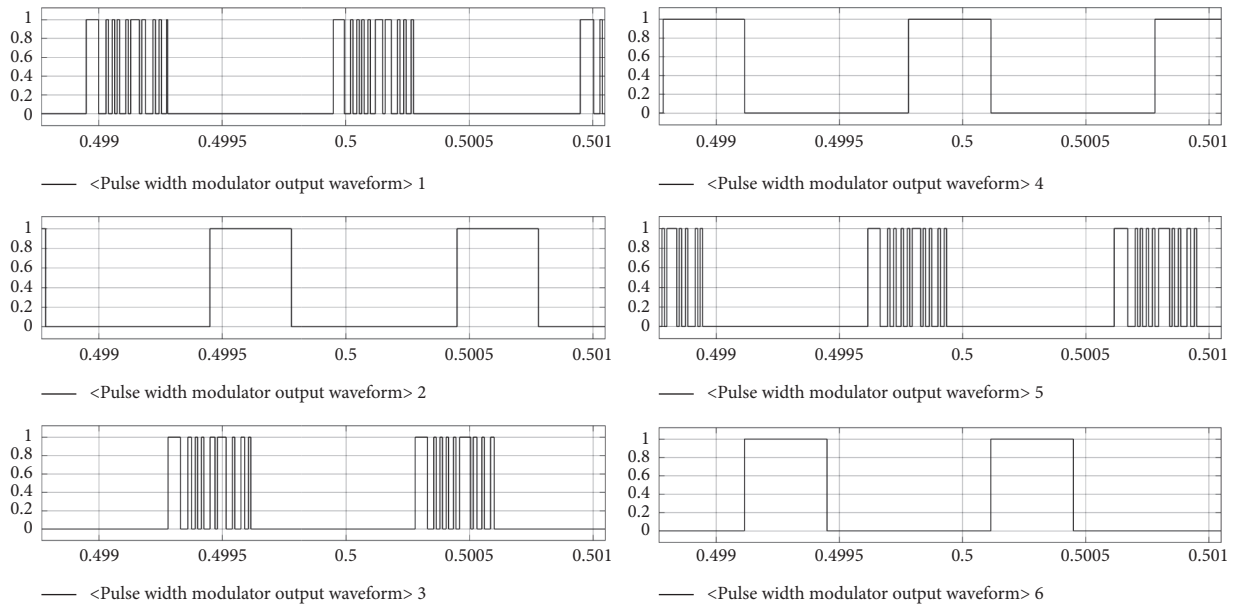


FIGURE 16: Pulse width modulator output waveform of the Homopolar motor of PI control.

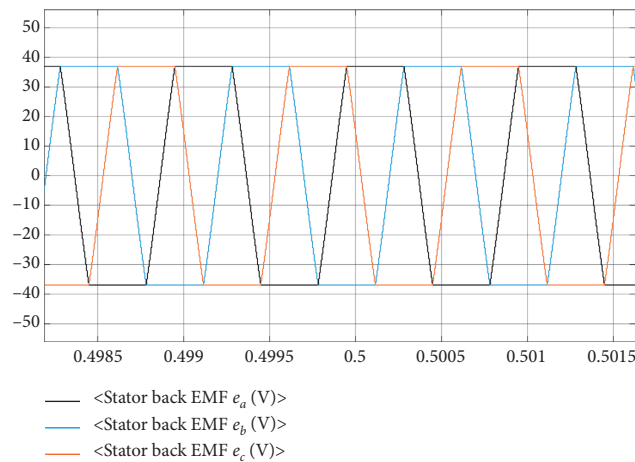


FIGURE 17: Three-phase-induced electromotive force waveform of the homopolar motor of PI control.



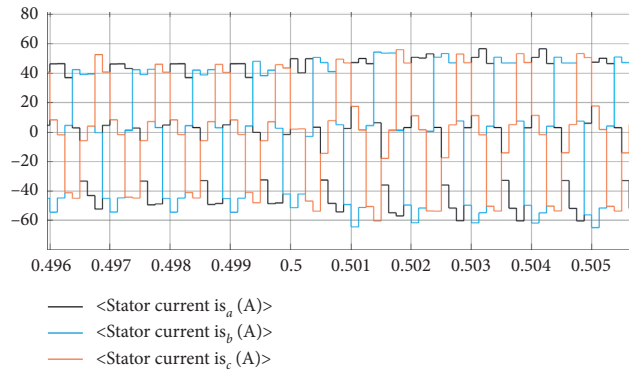


FIGURE 18: Three-phase-induced current waveform of the homopolar motor of PI control.

loading at 0.5 S, the state of the power switching tube of the PWM modulator changes and causes the current pulse amplitude to vary slightly.

It is thus clear that the simulation result basically coincides with the expected result, which suggests that the control method is feasible. The purpose of this research, however, is only to verify the correctness and feasibility of this control method by simulation and analysis considering the large inertia of flywheel and long acceleration time in the practical control of flywheel motor acceleration. The experiment result and the discussion part of work apply to general parameters and to the parameters used in experiment.

## 7. Conclusion

Flywheel energy storage will be a future direction for the field of energy saving, though traditional flywheel motors have the defects of great iron loss in rotation, poor rotor strength, and robustness. In view of these problems, the electrically excited homopolar motor is adopted for flywheel energy storage in this research for its advantages such as simple structure, high rotor strength, and low iron loss in high-speed rotation. A PI-controlled double closed-loop governing system of the homopolar motor was designed on simulation platform Simulink. And the charging process of the flywheel energy storage is studied in this paper. The simulation result shows that the control system of the electrically excited homopolar motor realized the regulating of motor phase current and torque by modulating PWM duty cycle in Hpwm-Lon mode with a speed-current controller using PID control strategy, realized fast speed governing, showed good static and dynamic performances, and achieved the purpose of quick charge by flywheel energy storage. This research is just a small part of the flywheel energy storage project which will continue the research and verification of the energy storage and discharge process later, use advanced motor intelligent control algorithm to realize homopolar motor's robust control, reduce the switching loss and torque ripple, and further improve the system performance.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Disclosure

We confirm that the content of the manuscript has not been published or submitted for publication elsewhere.

## Conflicts of Interest

The authors declare that there are no conflicts of interest.

## Authors' Contributions

All authors have read and approved the final version of the manuscript.

## Acknowledgments

We would like to thank my mentor Professor Dong for stimulating discussions with respect to the topic of this paper and laboratory equipment. This research was supported by Inner Mongolia Autonomous Region Natural Science Foundation (2017MS0521) and Research Program of Science and Technology at Universities of Inner Mongolia Autonomous Region (NJZY17352 and NJZY18232).

## References

- [1] S. Koohi-Fayegh and M. A. Rosen, "A review of energy storage types, applications and recent developments," *Journal of Energy Storage*, vol. 27, p. 101047, 2020.
- [2] B. H. Kenny, P. E. Kascak, R. Jansen, T. Dever, and W. Santiago, "Control of a high-speed flywheel system for energy storage in space applications," *IEEE Transactions on Industry Applications*, vol. 41, no. 4, pp. 1029–1038, 2005.
- [3] I. Alan and T. A. Lipo, "Induction machine based flywheel energy storage system," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 1, pp. 151–163, 2003.
- [4] Z. Wang and X. Xue, "A Type of electrically-excited homopolar rotary motor," pp. 4–3, Beijing, China, 2018, Patent no: CN107872104A.
- [5] Z. Lou, K. Yu, L. Wang, Z.-A. Ren, and C. Ye, "Two-reaction theory of homopolar inductor alternator," *IEEE Transactions on Energy Conversion*, vol. 25, no. 3, pp. 677–679, 2010.
- [6] Q. Xin, K. Yu, Z. A. Ren, Z. lou, and C. ye, "Inductance mathematic model of a homopolar inductor alternator in a novel pulse capacitor charge power supply," *IEEE Transactions on Plasma Science*, vol. 41, no. 5, pp. 1231–1236, 2013.

- [7] D. Stephen, X. Liu, S. Su et al., *Electric Machinery*, Publishing House of Electronics Industry, Beijing, China, 7th edition, 2016.
- [8] K. J. Astrom and T. Hagglund, *PID controllers : Theory Design and Tuning*, Instrument Society of America, Research Triangle Park, NC, USA, 2nd edition, 1995.
- [9] P.-D. Yang, *PID Controller Parameters Tuning Method and its Application*, China Electric Power Press, Beijing, China, 2016.
- [10] Z. He, *Parameters Tuning of PID Controller and It's Application*, Zhejiang University, Zhejiang, China, 2005.
- [11] A. Visioli and Q. Zhong, *Smith-principle-based PID-type Control*, Springer, Berlin, Germany, 2011.
- [12] I. Boiko, *Discontinuous Control Systems*, Birkhuser, Cambridge, MA, USA, 2009.
- [13] S. Manesis and N. George, *Introduction to Industrial Automation*, Taylor and Francis, Milton Park, UK, 2018.
- [14] G. P. Liu, *Control Systems Design*, Springer, Berlin Germany, 2005.
- [15] Birkhuser, *The Servo Problem in Discontinuous Control systems*, Birkhuser, Cambridge, MA, USA, 2009.
- [16] L. Ibarra and C. Webb, *Advantages of Fuzzy Control While Dealing With Complex/Unknown Model Dynamics: A Quadcopter Example*, IntechOpen, London, UK, 2016.
- [17] P. Tang, *Research on Magnetic Suspension Flywheel Electric Machine for Energy Storage and Control of Drive Systeem*, Harbin Institute of Technology, Harbin, China, 2009.
- [18] E. Sghaier, A. Bourdon, D. Remond, J.-L. Dion, and N. Peyret, *Dynamic Behavior of Very-High Speed Rotors at Non-stationary Conditions*, Springer International Publishing, Manhattan, NY, USA, 2018.
- [19] L. Jing, X. Xue, and N. Yang, "Research on fuzzy PID control strategy for double closed-loop control system of brushless DC motor," *Revista de la Facultad de Ingenieria*, vol. 9, pp. 8–14, 2017.
- [20] L. Jing and N. Yang, "Research on Switching PID control strategy for brushless direct current motor," *Boletin Tecnico/ Technlcal Bulletin*, vol. 55, pp. 676–683, 2017.
- [21] Q. Li, H. Huang, and B. Yin, "The study of PWM methods in permanent magnet brushless DC motor speed control system," in *Proceedings of the International Conference on Electrical Machines and Systems*, pp. 3897–3900, Wuhan, China, October 2008.

## Research Article

# DTFA-Net: Dynamic and Texture Features Fusion Attention Network for Face Antispoofing

Xin Cheng <sup>1</sup>, Hongfei Wang <sup>1</sup>, Jingmei Zhou <sup>2</sup>, Hui Chang <sup>1</sup>, Xiangmo Zhao <sup>1</sup>,  
and Yilin Jia<sup>3,4</sup>

<sup>1</sup>School of Information Engineering, Chang'an University, Xi'an 710064, China

<sup>2</sup>School of Electronic and Control Engineering, Chang'an University, Xi'an 710064, China

<sup>3</sup>Xi'an University of Architecture and Technology, Xi'an, China

<sup>4</sup>University of South Australia An De College, Adelaide, Australia

Correspondence should be addressed to Xin Cheng; [xincheng@chd.edu.cn](mailto:xincheng@chd.edu.cn), Hongfei Wang; [herr\\_wang@chd.edu.cn](mailto:herr_wang@chd.edu.cn), and Jingmei Zhou; [jmzhou@chd.edu.cn](mailto:jmzhou@chd.edu.cn)

Received 18 May 2020; Accepted 16 June 2020; Published 10 July 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Xin Cheng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

For face recognition systems, liveness detection can effectively avoid illegal fraud and improve the safety of face recognition systems. Common face attacks include photo printing and video replay attacks. This paper studied the differences between photos, videos, and real faces in static texture and motion information and proposed a living detection structure based on feature fusion and attention mechanism, Dynamic and Texture Fusion Attention Network (DTFA-Net). We proposed a dynamic information fusion structure of an interchannel attention block to fuse the magnitude and direction of optical flow to extract facial motion features. In addition, for the face detection failure of HOG algorithm under complex illumination, we proposed an improved Gamma image preprocessing algorithm, which effectively improved the face detection ability. We conducted experiments on the CASIA-MFSD and Replay Attack Databases. According to experiments, the DTFA-Net proposed in this paper achieved 6.9% EER on CASIA and 2.2% HTER on Replay Attack that was comparable to other methods.

## 1. Introduction

With the application of face recognition technology in the identification scene such as access security check and face payment, the methods of attack and fraud against face recognition system also appear. Face is obviously a much easier way to steal identity information than biometric features such as iris and fingerprints. Attackers can easily steal images or videos of legitimate users on social networking sites and then launching print or replay attacks on face recognition systems. Some face verification systems use techniques such as face tracking to locate key points on the face, requiring users to complete actions such as blinking, shaking their heads, and reading text aloud and use motion detection to determine whether the current image is a real face. This approach is not suitable for silent detection scenarios. In addition, some researchers use infrared camera,

depth camera, and other sensors to collect different modes of face images to achieve living detection [1–3]. These methods show excellent performance in many scenarios but need to add information acquisition equipment other than camera to face recognition devices, need to invest additional hardware costs, and cannot meet the requirements of some mobile devices. In this paper, we will study the monocular static and silent living detection and achieve the living detection task by analyzing the difference between real face and fake face in image texture, facial structure, action change, and so on.

Real face image is often taken directly by the camera, while attacking face images are collected many times. As shown in Figure 1, false face images may show the texture of the image carrier itself, and the light region with large difference from the real face image is also easy to appear in the false face image. According to this, researchers proposed

many feature descriptors for characterizing the living texture of face and then implemented the classification by training models such as SVM and LDA classifier. In order to characterize the high semantic features of face living body, the deep neural network is applied in the feature extraction process to further enhance the performance of living detection. The features included in the local area of the face can often be used as an important basis for living detection and play a different role, as shown in Figure 2. Based on this, some researchers [4, 5] decomposed faces into different regions to extract features through neural networks and then realize feature splicing.

Most prosthetic faces are difficult to simulate the vital signs of real faces, such as head movement, lip peristalsis, and blinking. At the same time, due to background noise, skin texture, and other factors, the dynamic characteristics of real face in some frequency bands are obviously higher than that of fraudulent face, which provides the basis for distinguishing real face from fraudulent face. The variation in optical flow field is an important basis of this kind of algorithm. However, the dynamic information generated by movement and bending of photo will influence the extraction of life signals. Remote photoplethysmography (rPPG) is another effective noncontact living signal extraction method, which provides a basis for face living detection by observing face images to calculate the changes in blood flow and flow rate [6, 7], but the rPPG method has strict requirements for algorithm application environment.

This work proposed a network that fuses dynamic and texture information to represent face and detect the attacks. Optical flow method is used to calculate the motion change in two adjacent frames of face images. The optical flow generated by the bending and movement of the photo is different from the optical flow generated by the movement of the real face in the direction of displacement. We use a simple convolutional neural network with the same structure to characterize the magnitude and direction of displacement. Then, a feature fusion module is designed for the combination of the above two representations so that, on this basis, facial motion features can be further extracted. In addition, RGB images are used to extract texture information of the face area. By giving a different attention to the parts of the face, we enhance the network's ability to represent living faces.

Face detection algorithms are widely used in living body detection tasks, which can be used to locate faces, thereby eliminating the interference of background information on living body detection. In this paper, for face detection scenes under complex lighting, we propose an improved image preprocessing algorithm combined with local contrast in the face area, which effectively improves the performance of the face detection algorithm.

## 2. Relating Works

*2.1. Texture based.* Living verification is completed by using the difference between real face and replay image in surface texture, 3D structure, image quality, and so on. Boulkenafet et al. [8] analyzed the chroma and brightness difference

between real and false face images, it is based on the color local binary pattern, and the feature histogram of each order image frequency band was extracted as the face texture representation. Finally, the classification was realized by support vector machine, and testing on the Replay Attack Dataset obtained the half error rate; it is 2.9%. Galbally et al. [9] prove that the image quality loss value produced by Gaussian filtering can distinguish the truth effectively with fraudulent face images, designed a quality assessment vector containing 14 indicators, and proposed a live detection method, the method in combination with LDA (linear discriminant analysis), and obtained 15.2% half error rate on the Replay Attack Dataset. However, such methods based on static feature often require the design of specific descriptors for a certain types of attacks, and the robustness is poor under different light conditions and different fraud carriers [10].

*2.2. Dynamic Based.* Some researchers have proposed a face living detection algorithm based on dynamic features by analyzing face motion patterns and show good performance in related datasets [11]. Kim et al. [12] designed a local velocity pattern for the estimation of the speed of light and distinguished the fraud from the real face according to the difference in the diffusion speed between the light on the real face and the fraud carrier surface. A 12.50% half error rate was obtained on the Replay Attack Dataset. Bharadwaj et al. [13] amplify the blink signal which is 0.2–0.5 Hz in the image by the Eulerian motion amplification algorithm, combined with local binary pattern with directional flow histogram (LBP-HOOF) to extract dynamic features as classification basis and obtained error rate which is 1.25% on the Replay Attack Dataset. At the same time, they proved the positive effect of image amplification algorithm on the performance of the algorithm. Freitas et al. [14] learned from the facial expression detection method, extracted feature histograms from the orthogonal plane of time-spatial domain by using LBP-TOP operator, used support vector machine to classify, and got 7.6% half error rate on Replay Attack Dataset. Xiaoguang et al. [15] based on the action information between adjacent frames established a CNN-LSTM network model, used convolutional neural network to extract the texture features of adjacent frame face images, and then input it to the long- and short-term memory structure to learn the time-domain action information in face video.

In addition, some researchers combined different detection equipments or system modules to fuse information on different levels, which effectively increased the accuracy of living detection [1, 16]. Zhang and Wang [17] used Intel RealSense SR300 camera to construct multimodal face image database including RGB image, depth image (depth), and infrared image (IR). The face region was accurately located using face 3D reconstruction network PRNet [18] and mask operation and then based on ResNet 18 classification [19] network to extract and fuse feature of multimodal data which mixed RGB, depth, and IR.



FIGURE 1: Face print and replay attack images. The face attacked has been collected many times, showing the difference between texture feature, light, image quality, and real face.

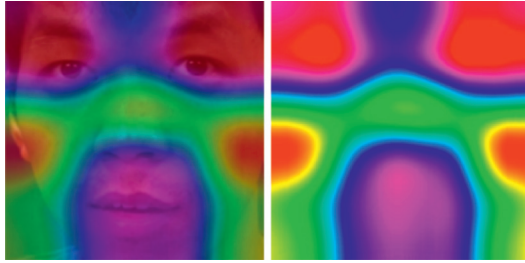


FIGURE 2: Weights visualization of a layer in a depth neural network for real face texture information extraction. Different face regions occupy different weights in living detection task.

### 3. Proposed Method

**3.1. Face Detection in Complex Illumination.** In order to eliminate the interference of background in the process of living information extraction, it is necessary to segment the face area of the image. Traditional detection techniques can be divided into three categories: the face detection based on feature, the face detection based on template, and the face detection based on statistics. This paper uses face front detection API provided by Dlib, which uses gradient direction histogram feature to achieve face detection. The face detection algorithm based on gradient direction histogram can maintain good immutability of image texture and optical deformation and ignore the slight texture and changes in expression.

Histogram of Oriented Gradients (HOGs) is a method used to describe the local texture features of image. The algorithm divides the image into small spaces and calculates the gradient of pixel points in each space. The pixel point gradient calculation is shown in the following equations:

$$G_x(x, y) = I(x + 1, y) - I(x - 1, y), \quad (1)$$

$$G_y(x, y) = I(x, y + 1) - I(x, y - 1), \quad (2)$$

where  $G_x(x, y)$  and  $G_y(x, y)$  are the horizontal gradient and vertical gradient at the  $(x, y)$  of the image, respectively,

and  $I(x, y)$  is the gray value. In reality, local shading or over exposure will affect the extraction of gradient information because the image target will appear in different light environments, as shown in Figure 3. In order to enhance the robustness of the HOG feature descriptor to environmental changes and reduce the noise such as the local shadow of the image, a Gamma correction algorithm is used to preprocess the image to eliminate the interference of partial light.

Traditional Gamma correction method changes the brightness of image by selecting the appropriate  $\gamma$  operator, as follows:

$$O(x, y) = 255 \times \left[ \frac{I(x, y)}{255} \right]^\gamma, \quad (3)$$

where  $I(x, y)$  is the pixel value of the image at the position  $(x, y)$ ,  $O(x, y)$  is the corrected pixel value, and  $\gamma$  is the constant. The traditional method performs image processing at the global level without considering the lightness difference between local and neighborhood pixels. Therefore, Schettini et al. [20] proposed a formula for the value of  $\gamma$  operator:

$$\gamma[x, y] = \partial^{[128 - \text{mask}(x, y)/128]},$$

$$\partial = \begin{cases} \frac{(\ln(\bar{I}/255))}{(\ln(0.5))}, & \bar{I} < 128, \\ 1, & \bar{I} = 128, \\ \frac{(\ln(0.5))}{(\ln(\bar{I}/255))}, & \bar{I} > 128, \end{cases} \quad (4)$$

where mask is an image mask and Gaussian blur can be used in practice. For the more balanced image with bright area and dark area, the average pixel of the image is close to 128, so the calculated  $\alpha$  is close to 1, and the image is hardly changed, which obviously does not meet the actual needs. Considering the local feature of face, this paper introduces

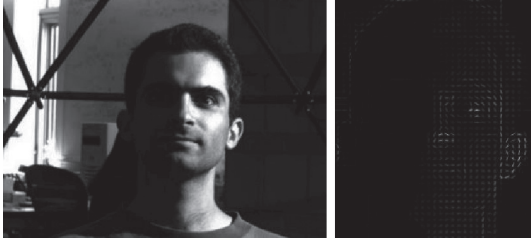


FIGURE 3: HOG feature of shadow on face region under light condition. It is necessary to initialize the face image because the shadow or exposure caused by complex light can affect the face region gradient information.

the local normalization method proposed in [21] to calculate the ratio relation of pixels in the neighborhood and adjust the operator  $\alpha$ :

$$\partial(x, y) = \begin{cases} \frac{(\ln(\bar{I}/255))}{(\ln(0.5))} + \frac{N(x, y)}{(\ln(\bar{I}/255))} \ln(0.5), & \bar{I} < 128, \\ \frac{(\ln(0.5))}{(\ln(\bar{I}/255))} + \frac{N(x, y)}{(\ln(0.5))} \ln\left(\frac{\bar{I}}{255}\right), & \bar{I} \geq 128. \end{cases} \quad (5)$$

Among them, the specific calculation process of local normalized characteristic  $N$  is as follows:

- (1) To calculate the maximum pixel value  $I_m(x, y)$  in the neighborhood  $\varphi(x, y)$  centered on pixel  $(x, y)$ ,

$$I_m(x, y) = \max\{I(i, j) \mid (i, j) \in \varphi(x, y)\}. \quad (6)$$

- (2) To calculate the median value of the  $I_m(x, y)$  of all pixels centered on pixel  $(x, y)$ ,

$$I_{mm}(x, y) = \text{medium}\{I_m(i, j) \mid (i, j) \in \varphi(x, y)\}. \quad (7)$$

- (3) To calculate the maximum value of the  $I_{mm}(x, y)$  of all pixels centered on pixel  $(x, y)$ ,

$$S(x, y) = \max\{I_{mm}(i, j) \mid (i, j) \in \varphi(x, y)\}. \quad (8)$$

- (4) To calculate the ratio of pixels  $(x, y)$  to neighborhood pixels,

$$N(x, y) = \frac{I(x, y)}{S(x, y)}. \quad (9)$$

We use algorithm in [20] and the improved algorithm in this paper to preprocess the portrait 208 photos on YaleB subdatabase that is difficult to be detected by HOG under complex lighting conditions and then detect 196 and 201 faces separately. The result is shown in Figure 4.

**3.2. DTFA-Net Architecture.** In Section 3.2, we mainly introduce the dynamic and texture features fusion attention network DTFA-Net. As shown in Figure 5, the optical flow graph and the texture image are, respectively, subjected to obtain  $256 \times 2$  and  $256 \times 4$  embedding by extracting dynamic feature and texture feature from subnetwork and then fusing the spliced  $256 \times 6$  features through the fully connected layer and living detection. The specific details of the network are described below.

**3.2.1. Dynamic Feature Fusion.** This paper generates the optical flow field change map of adjacent two frames of face video by the optical flow method. The optical flow change in face region is extracted by dynamic feature fusion subnetwork in two dimensions of displacement and size, and the features of the two dimensions are fused by feature fusion block to extract the dynamic information of face region.

(1) *Optical Flow.* Optical flow method is a proposal used to describe the motion information of adjacent frame objects. It reflects the interframe field changes by calculating the motion displacement in the  $x$  and  $y$  directions of the image on the time domain. Defining video midpoint  $P$  located  $(x, y)$  of the image at the  $t$  moment and moving to the place  $(x + dx, y + dy)$ , then when the  $dt$  is close to 0, the two pixel values satisfy the following relationship:

$$I(v) = I(v + d), \quad (10)$$

where  $v = (x, y)$  is the coordinate of the point  $P$  at the time  $t$ ,  $I(v)$  is the gray value of the place  $(x, y)$  at the time  $t$ ,  $d = (dx, dy)$  is the displacement of the point  $P$  during  $dt$ , and  $I(v + d)$  is the gray value of the place  $(x + dx, y + dy)$  at the time  $t + dt$ .

In this paper, the dense optical flow method proposed by Farneback [22] is used to calculate the interframe displacement of face video. The algorithm approximates the pixels of two-frame images by a polynomial expansion transformation. And it based on the assumption that the local optical flow and the image gradient are stable, and the displacement field is deduced in the polynomial expansion coefficient. We transform the displacement  $d = (dx, dy)$  to the extreme coordinate system  $\bar{d} = (\rho, \theta)$  and visualize the optical flow displacement and direction by the HSV model. As shown in Figure 6, the optical flow change image obtained will be used as input of the dynamic feature fusion network.

(2) *Fusion Attention Module.* In the process of dynamic information extraction, we extract, respectively, the motion information contained in the input optical flow change direction feature map and the optical flow change intensity feature map through 5 convolution layers. Because the motion pattern of living human face contains two dimensions of direction and intensity, it is necessary to combine the above representations to further extract the moving features of the face. As a result, we designed a fusion module, as shown in Figure 7.

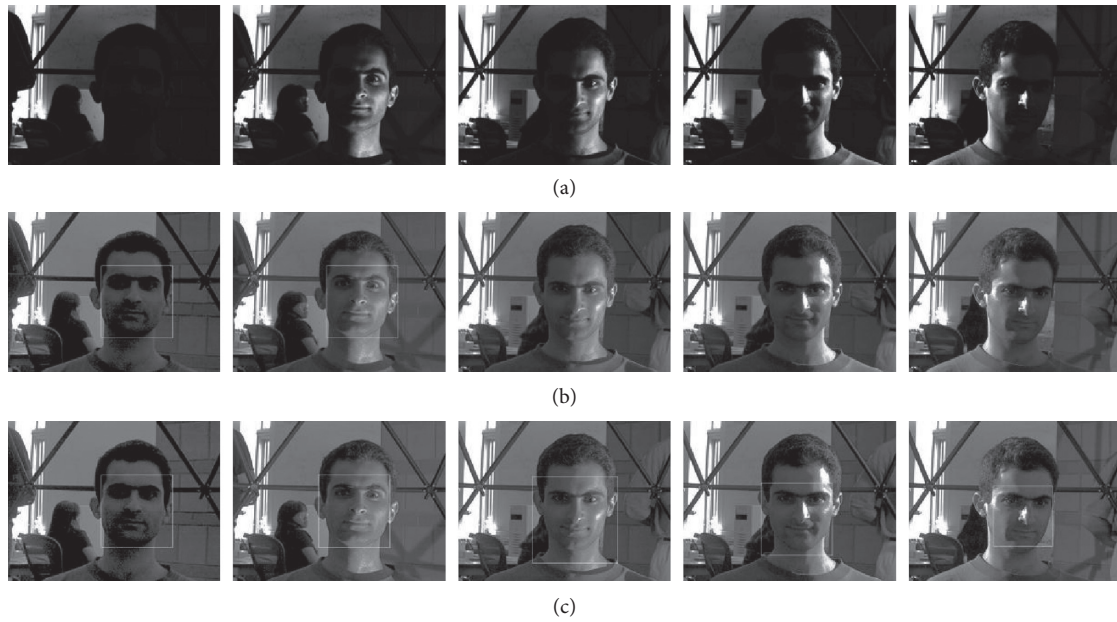


FIGURE 4: Comparison between [20] and ours. The improved algorithm we proposed performs better than that in [20]. (a) Original images that cannot detect face by HOG; (b) images processed in [20] and the detection result; (c) images processed by ours and the detection result.

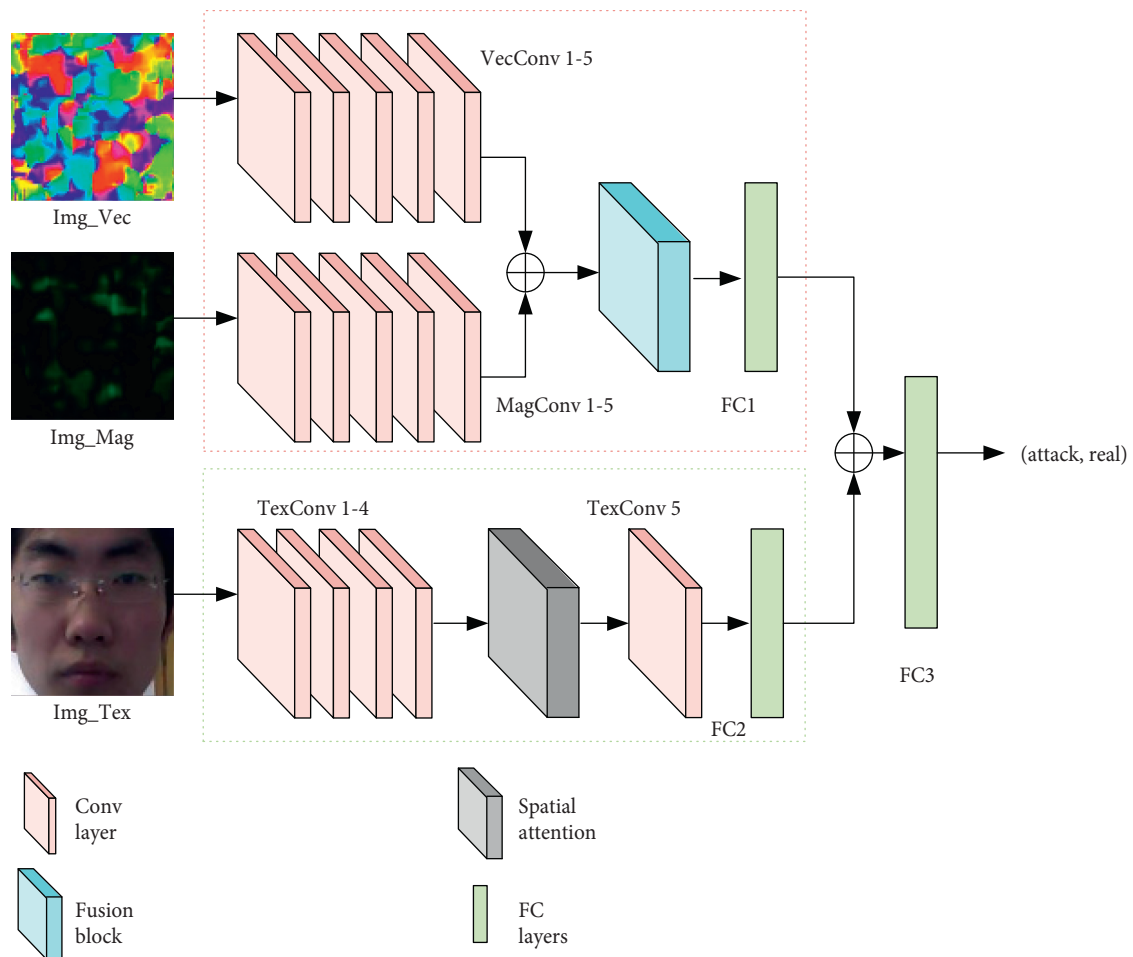


FIGURE 5: The dynamic and texture features fusion attention network (DTFA-Net) architecture. The figure omits the ReLU and pooling layers after the convolution layer, and the id of the convolution is shown on the top. Color code used is as follows: pink = convolution; blue = fusion block; gray = spatial attention; green = fully connected layer.

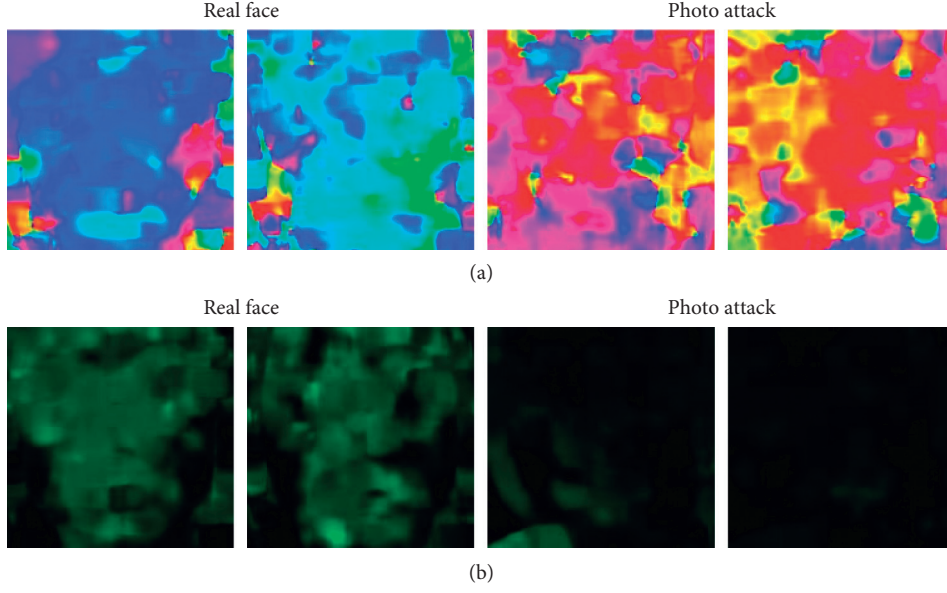


FIGURE 6: Optical flow visualization of two adjacent face regions: (a) visualization of changes in optical flow direction: hue = direction of optical flow, saturation = 255, and value = 255; (b) optical flow magnitude visualize: hue = 255, saturation = 255, and value = size of optical flow. Among them, the left two are the optical flow changes in the real face, and the right two groups are the optical flow changes in the photo attacks.

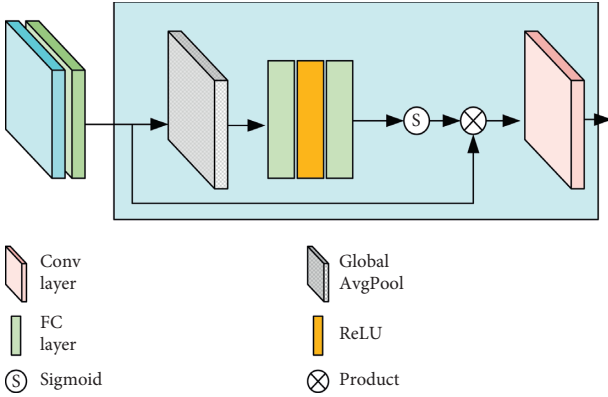


FIGURE 7: Fusion attention module architecture.

To improve the characterization ability of the model, we use the SE structure [23] in the fusion module, which gives different weights for the optical flow intensity, and direction features to strengthen the decision-making ability of some features. First, global pooling of feature graphs is

$$\text{op}_c = \text{AvgPool}(F_{\text{op}}) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F_{\text{op}}(i, j), \quad (11)$$

where  $F_{\text{op}}(i, j)$  stands for the concatenated features of optical magnitude and angle. Through global average pooling, the dimension of the stitching feature map changes from  $C \times H \times W$  to  $C \times 1 \times 1$ . Secondly, learn the nonlinear functional relationship between each channel through full connection (FC) and activation function (ReLU). Then, use normalization (sigmoid) to get the weight of each channel:

$$\text{op}_a = \sigma(\text{FC}(\delta(\text{FC}(\text{op}_c))))), \quad (12)$$

where  $\sigma$  is the sigmoid function and  $\delta$  is the ReLU function. The two fully connected layers are used to reduce and recovery dimension, respectively, which is helpful to improve the complexity of the function. Finally, we multiply  $F_{\text{op}}$  with  $\text{op}_a$  and pass through a convolution layer to get the fusion features:

$$F_{\text{op}^+} = \text{Conv}(\text{op}_a \otimes F_{\text{op}}). \quad (13)$$

(3) *Network Details.* Dynamic feature extraction subnetwork input image size is  $227 \times 227 \times 3$ , which contains 11 convolution layers, 2 full connected layers, and 6 pooling layers. Tables 1–3 show the specific network parameters of convolution and pooling layers.

3.2.2. *Texture Feature Representation.* In specific, we map the input RGB image to the intermediate feature maps with a dimension of 384 through TexConv1-4 and then pay more attention to some of the regions through the spatial attention mechanism and then input the output of the attention module to TexConv5 and full connection layer FC2 performs feature extraction. The structure of the convolutional layer TexConv1-5 is shown in Table 1, and the structure of the fully connected layer FC2 is shown in Table 4.

(1) *Spatial Attention Block.* After experiments, we found that neural networks often pay special attention to the human eyes, cheeks, mouths, and other areas when extracting living features. Therefore, we added a spatial attention module to the static texture extraction structure and give a different



TABLE 1: The network of Mag\_Conv1-5.

| Layer     | Input size    | Kernel size | Filter | Stride |
|-----------|---------------|-------------|--------|--------|
| Mag_Conv1 | 227 * 22 * 3  | 11 * 11 * 3 | 96     | 4      |
| MaxPool1  | 27 * 27 * 64  | 3 * 3       |        | 2      |
| Mag_Conv2 | 27 * 27 * 64  | 5 * 5 * 64  | 192    | 1      |
| MaxPool2  | 27 * 27 * 192 | 3 * 3       |        | 2      |
| Mag_Conv3 | 13 * 13 * 192 | 3 * 3 * 192 | 384    | 1      |
| Mag_Conv4 | 13 * 13 * 384 | 3 * 3 * 384 | 256    | 1      |
| Mag_Conv5 | 13 * 13 * 256 | 3 * 3 * 256 | 256    | 1      |

\*VecConv1-5 and TexConv1-5 parameters are same as MagConv1-5.

TABLE 2: The structure of fusion attention module.

| Layer         | Input size    | Kernel size | Filter | Stride |
|---------------|---------------|-------------|--------|--------|
| GlobalAvgPool | 13 * 13 * 512 | 13 * 13     |        | 1      |
| Fc1           | 512           |             |        |        |
| Fc2           | 32            |             |        |        |
| LK_Conv       | 13 * 13 * 512 | 3 * 3 * 512 | 256    | 2      |
| LK_MaxPool    | 6 * 6 * 256   | 3 * 3       |        | 1      |

TABLE 3: The structure of FC1 in Figure 5.

| Layer | Input size  | Output size |
|-------|-------------|-------------|
| FC1_1 | 256 * 6 * 6 | 256 * 3 * 3 |
| FC1_2 | 256 * 3 * 3 | 256 * 2 * 2 |
| FC1_3 | 256 * 2 * 2 | 256 * 2     |

TABLE 4: The structure of FC2-3 in Figure 5.

| Layer | Input size  | Output size |
|-------|-------------|-------------|
| FC2_1 | 256 * 6 * 6 | 256 * 3 * 3 |
| FC2_2 | 256 * 3 * 3 | 256 * 2 * 2 |
| FC3_1 | 256 * 6     | 256 * 3     |
| FC3_2 | 256 * 3     | 256         |
| FC3_3 | 256         | 2           |

attention to the features of different face regions. We adopted the CBAM (Figure 8) spatial attention structure proposed in [24]. This module reduces the dimension of the input feature map through the maximum pooling and average pooling layers, splices the two feature maps, and obtains the attention weight of  $1 * H * W$  by the convolution layer and activation function:

$$SA_c = \delta(\text{Conv}(\text{Cat}(\text{AvgPool}(F_t), \text{MaxPool}(F_t)))). \quad (14)$$

Finally, we utilized element-wise product for input  $F_t$  and  $SA_c$ , and the output of the spatial attention block will pass through the next layers, TextConv5 and FC2:

$$F_{t+} = SA_c \otimes F_t. \quad (15)$$

**3.2.3. Feature Fusion.** Through the above two subnetworks, dynamic information and texture information are obtained, respectively. By a series of fully connected layers, dropout layers, and activation functions, we fully fuse the two

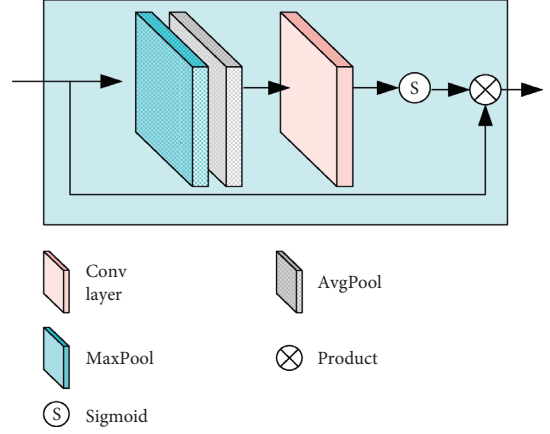


FIGURE 8: Spatial attention block. We introduce this module after the convolution layer of the subnetwork is extracted from the static feature, which gives the difference attention to the local area of the face.

information, learning the nonlinear relationship between the dynamic and static features, and obtain a two-dimensional representation of face in living information for living detection, as shown in Table 4.

## 4. Experiment

**4.1. Dataset.** We use CASIA-MFSD [25] to train and test the model. The dataset contains a total of 600 face videos collected from 50 individuals. Face video of real face, photo attack, and video attack scenes are collected at different resolutions. Among them, photo attack includes photo bending and photo mask. We ignore the different attack ways and divide all the videos into real face and false face. Through the calculation of optical flow field, face region detection and tailoring, etc., get 35428 sets of training images and 64674 sets of test images, as shown in Figure 9. And we also train and test our model on Replay Attack Database.

**4.2. Evaluation.** This experiment uses false acceptance rate (FAR), false rejection rate (FRR), equal error rate (EER), and half total error rate (HTER). The face living detection algorithm is based on these indicators. The FAR refers to the ratio of judging the fake face as the real face; the FRR refers to the ratio of judging the real face as false, and the calculation formulas are shown as follows:

$$FAR = \frac{N_{f_r}}{N_f}, \quad (16)$$

$$FRR = \frac{N_{r_f}}{N_r}, \quad (17)$$

where  $N_{f_r}$  is the number of false face error,  $N_{r_f}$  is the number of real face error,  $N_f$  is the number of false face liveness detection, and  $N_r$  is the number of real face detection. The two classification methods of this experiment are as follows: (1) nearest neighborhood (NN), which corresponds the two-dimensional vector, of which each

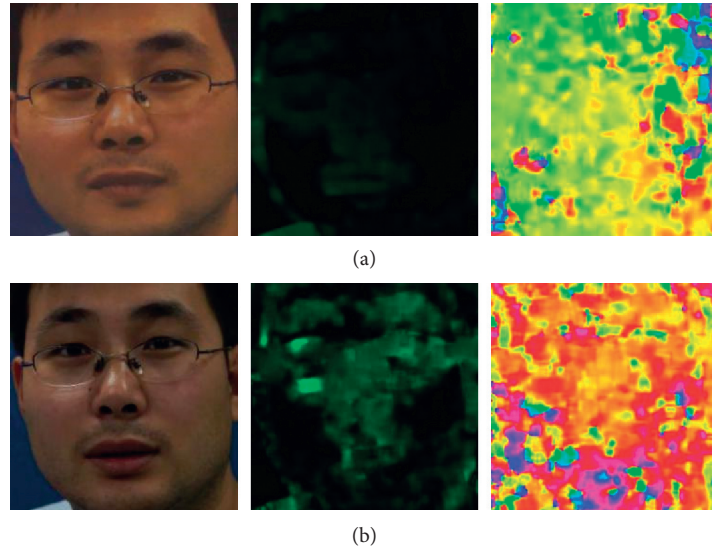


FIGURE 9: CASIA-MFSD examples after preprocessing. From left to right: texture image, optical flow magnitude, and optical flow direction. Among them, (a) fake face and (b) real face.

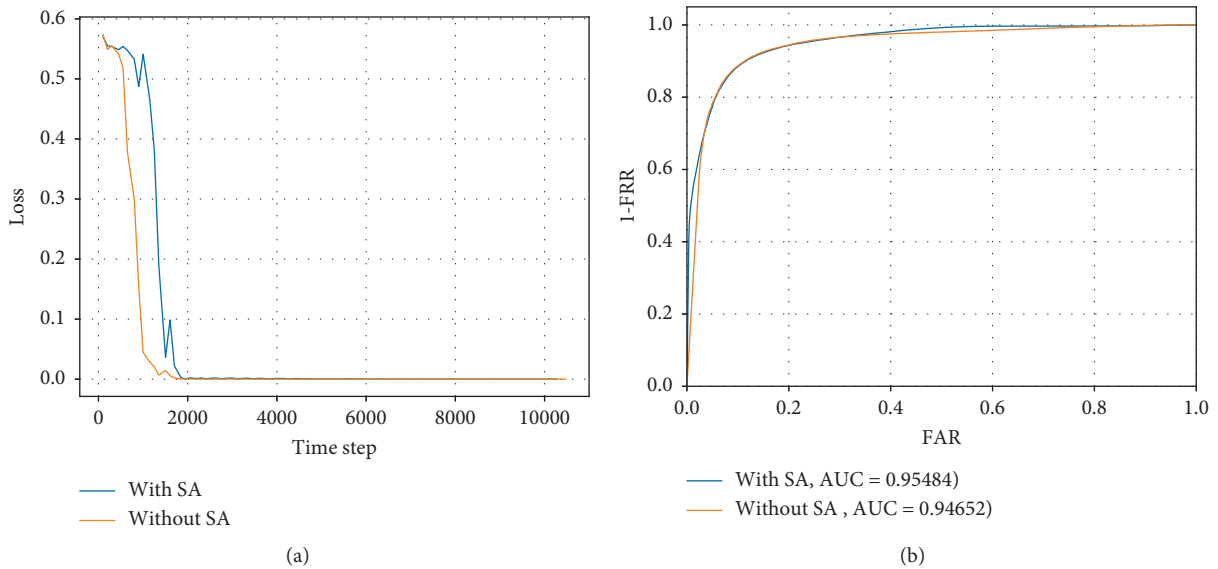


FIGURE 10: Network loss and ROC curve with or without spatial attention module: (a) training loss as time step went by; (b) ROC curve.

dimension value represents the probability of real face or attack face and selects the category which corresponds to the maximum value as the classification result. (2) Thresholding selects a certain threshold to classify the representation result. This method is mainly for model validation and testing. Calculating FAR and FRR at different thresholds can plot the receiver operating characteristic (ROC) curve for measuring the nonequilibrium in the classification problem; the area under the ROC curve (area under curve, AUC) can intuitively show the algorithm classification effect.

**4.3. Implementation Details.** The proposed method is implemented in Pytorch with an inconstant learning rate (e.g.,  $lr = 0.01$  when  $epoch < 5$  and  $lr = 0.001$  when  $epoch \geq 5$ ).

The batch size of the model is 128 with  $num\_worker = 100$ . We initialize our network by using the parameters of AlexNet100. The network is trained with standard SGD for 50 or 100 epochs on Tesla V100 GPU. And we use cross entropy loss, and the input resolution is  $227 \times 227$ .

#### 4.4. Experimental Result

**4.4.1. Ablation of Spatial Attention Module.** We conducted an ablation experiment on the attention module of the texture feature extraction subnetwork and only rely on texture features to perform live detection on the CAISA dataset. We trained the two texture feature extraction networks with or without spatial attention block 50 times,

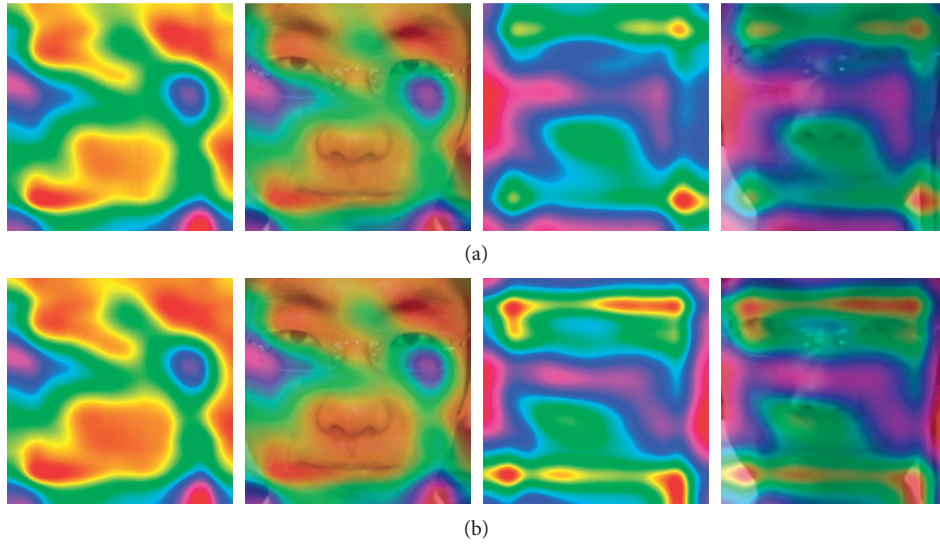


FIGURE 11: Change in weight heat map before and after spatial attention module: (a) before; (b) after. The spatial attention module pays special attention to some features of the face area.

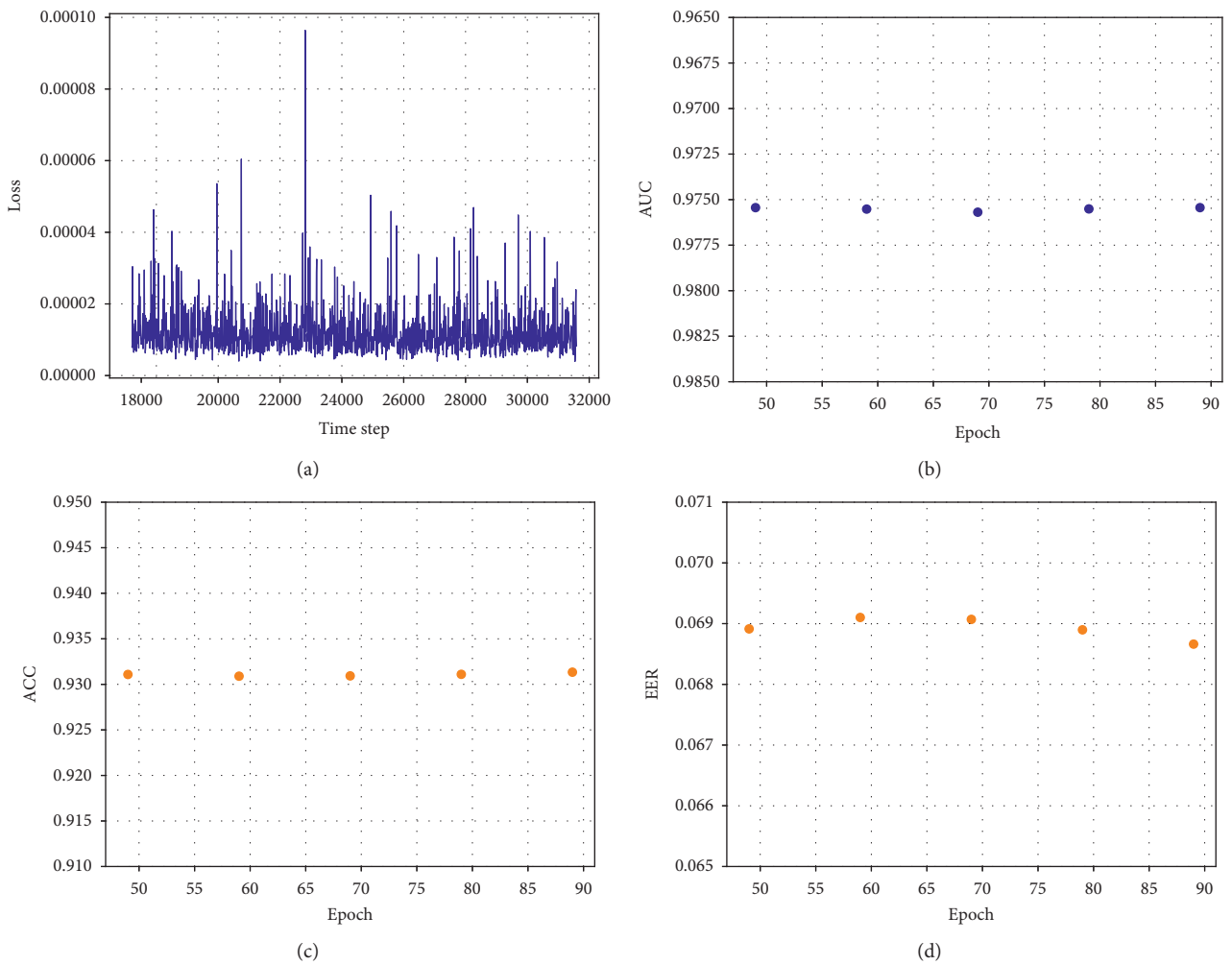


FIGURE 12: DTFA-Net training and evaluation results in Epoch49-89: (a) the loss fluctuations of model training in Epoch49-89; (b) the AUC results of the model in the test set in Epoch49-89; (c) the ACC results of the model in the test set in Epoch49-89; (d) the ERR results of the model in the test set in Epoch49-89.

TABLE 5: Comparison between our proposed method and the other in intradatabase.

| Method          | CASIA-MFSD |  | Replay Attack |          |
|-----------------|------------|--|---------------|----------|
|                 | EER (%)    |  | EER (%)       | HTER (%) |
| LBP [26]        | 18.2       |  | 13.9          | 13.8     |
| IQA [9]         | 32.4       |  | –             | 15.2     |
| CNN [4]         | 7.4        |  | 6.1           | 2.1      |
| LiveNet [27]    | 4.59       |  | –             | 5.74     |
| DTFA-Net (ours) | 6.90       |  | 6.47          | 2.2      |

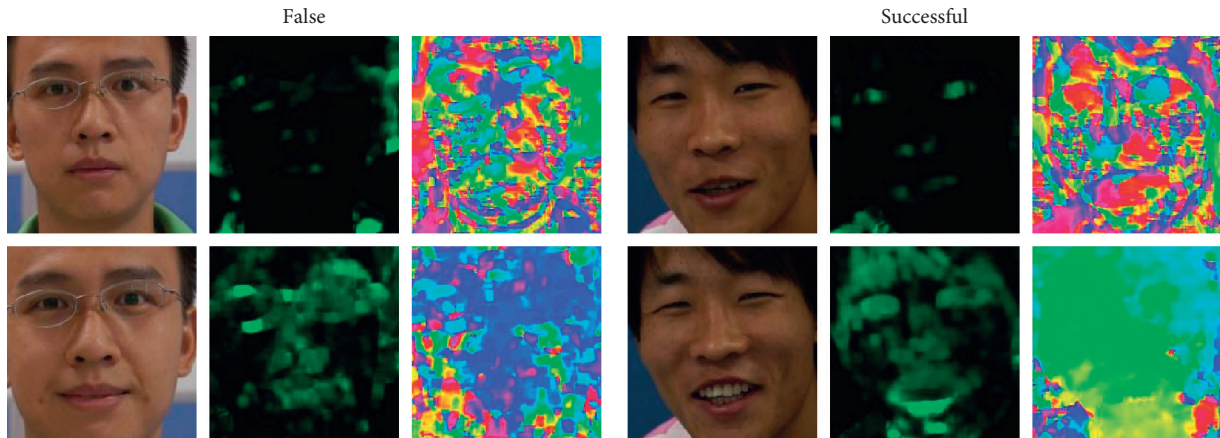


FIGURE 13: The false and right detection samples. Left: false-negative result; right: true-positive case.

respectively, and verified them on the CASIA test set. Figure 10 shows the training loss process (Epoch0-Epoch29) and the ROC curve in the test set (Epoch50). The experiment shows that, after introducing the attention mechanism, due to the increase in the network structure (in fact, a convolution layer is added), the loss of the model during the training process is slower than that of model without SA in the initial stage of training and there is a large shock. However, as the number of network training iterations increases, the loss tends to be stable, and there is almost no difference between the two cases. After 50 cycles of training, the model with SA achieved  $AUC = 95.4\%$  on the test set, which is higher than model without SA.

Visualize the input and output results of our spatial attention mechanism module, as shown in Figure 11. It shows that SA pays more attention to local areas in the face image, such as the mouth and eyes. This point shows the consistency of the prior knowledge as assumed by the traditional image feature description method.

We first do not use SA to train the DTFA network to a certain degree and then add the SA structure to train 100 times so that the spatial attention module can better learn face area information and accelerate model convergence. Figure 12 shows the training and test results of DTFA-Net on the CASIA dataset. When the number of training iterations of the model reaches the interval of 49–89,  $EER = 0.069$  and  $AUC = 0.975 \pm 0.0001$ , reaching a stable state.

Table 5 provides a comparison between the results of our proposed approach and those of the other methods in both

intradatabase evaluation. Our model result is comparable to the state-of-the-art methods.

*4.5. Samples.* Figure 13 shows several samples of the failure and right detection of real faces. Through analysis, we found that the illumination in RGB images may be the main cause of wrong classification.

## 5. Conclusion

This paper analyzed the photo and video replay attacks of face spoofing and built an attention network structure that integrated dynamic-texture features and designed a dynamic information fusion module that extracted features from texture images based on the spatial attention mechanism. At the same time, an improved gamma image optimization algorithm was proposed for preprocessing of image in face detection tasks under multiple illuminations.

## Data Availability

The CASIA-MFSD data used to support the findings of this study were supplied by CASIA under license and so cannot be made freely available. Requests for access to these data should be made to CASIA via <http://www.cbsr.ia.ac.cn>.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments


This work was supported by the National Key Research and Development Program of China (Grant 2018YFB1600600), National Natural Science Funds of China (Grant 51278058), 111 Project on Information of Vehicle-Infrastructure Sensing and ITS (Grant B14043), Shaanxi Natural Science Basic Research Program (Grant nos. 2019NY-163 and 2020GY-018), Joint Laboratory for Internet of Vehicles, Ministry of Education-China Mobile Communications Corporation (Grant 213024170015), and Special Fund for Basic Scientific Research of Central Colleges, Chang'an University, China (Grant nos. 300102329101 and 300102249101).

## References

- [1] H. Steiner, A. Kolb, and N. Jung, "Reliable face anti-spoofing using multispectral swir imaging," in *Proceedings of the International Conference on Biometrics*, IEEE, Halmstad, Sweden, May 2016.
- [2] Y. H. Tang and L. M. Chen, "3d facial geometric attributes based anti-spoofing approach against mask attacks," in *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, IEEE, Washington, DC, USA, pp. 589–595, September 2017.
- [3] R. Raghavendra and C. Busch, "Novel presentation attack detection algorithm for face recognition system: Application to 3d face mask attack," in *Proceedings of the IEEE International Conference on Image Processing*, IEEE, Paris, France, pp. 323–327, October 2014.
- [4] J. W. Yang, Z. Lei, and S. Z. Li, "Learn convolutional neural network for face anti-spoofing," 2014, <http://arxiv.org/abs/1408.5601>.
- [5] Y. Atoum, Y. J. Liu, A. Jourabloo, and X. M. Liu, "Face antispoofing using patch and depth-based cnns," in *Proceedings of IEEE International Joint Conference on Biometrics*, IEEE, Denver, Colorado, USA, pp. 319–328, August 2017.
- [6] J. Hernandez-Ortega, J. Fierrez, A. Morales, and P. Tome, "Time analysis of pulse-based face anti-spoofing in visible and nir," in *Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops*, IEEE, Salt Lake City, Utah, USA, June 2018.
- [7] S. Q. Liu, X. Y. Lan, and P. C. Yuen, "Remote photoplethysmography correspondence feature for 3d mask face presentation attack detection," in *Proceedings of the European Conference on Computer Vision*, IEEE, Munich, Germany, pp. 558–573, September 2018.
- [8] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face anti-spoofing based on color texture analysis," in *Proceedings of the International Conference on Image Processing*, IEEE, Quebec, Canada, pp. 2636–2640, September 2015.
- [9] J. Galbally and S. Marcel, "Face anti-spoofing based on general image quality assessment," in *Proceedings of the International Conference on Pattern Recognition*, IEEE, Stockholm, Sweden, pp. 1173–1178, August 2014.
- [10] Z. Boulkenafet, J. Komulainen, and A. Hadid, "On the generalization of color texture-based face anti-spoofing," *Image and Vision Computing*, vol. 77, pp. 1–9, 2018.
- [11] S. Tirunagari, N. Poh, D. Windridge, A. Iorliam, N. Suki, and A. T. S. Ho, "Detection of face spoofing using visual dynamics," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 762–777, 2015.
- [12] W. Kim, S. Suh, and J.-J. Han, "Face liveness detection from a single image via diffusion speed model," *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2456–2465, 2015.
- [13] S. Bharadwaj, T. Dhamecha, M. Vatsa et al., "Computationally efficient face spoofing detection with motion magnification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Portland, Oregon, June 2013.
- [14] T. Freitas, J. Komulainen, Anjos et al., "Face liveness detection using dynamic texture," *EURASIP Journal on Image and Video Processing*, vol. 2014, no. 1, p. 2, 2014.
- [15] T. U. Xiaoguang, H. Zhang, X. I. E. Mei et al., "Enhance the motion cues for face anti-spoofing using cnn-lstm architecture," 2019, <http://arxiv.org/abs/1901.05635>.
- [16] A. Alotaibi and A. Mahmood, "Deep face liveness detection based on nonlinear diffusion using convolution neural network," *Signal, Image and Video Processing*, vol. 11, no. 4, pp. 713–720, 2017.
- [17] S. Zhang and X. Wang, "A dataset and benchmark for large scale multi modal face anti-spoofing," in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, IEEE, CA, USA, November 2019.
- [18] Y. A. O. Feng, W. U. Fan, S. H. A. O. Xiaohu et al., "Joint 3D face reconstruction and dense alignment with position map regression network," in *Proceedings of the European Conference on Computer Vision*, Springer, Berlin, Germany, pp. 557–574, September 2018.
- [19] H. Kaiming, Z. Xiangyu, and R. Shaoqing, "Deep residual learning for image recognition," in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, June 2016.
- [20] Schettini, R. Gasparini, F. Corchs et al., "Contrast image correction method," *Journal of Electronic Imaging*, vol. 19, no. 2, Article ID 023005, 2010.
- [21] Y. Cheng, L. Jiao, X. Cao, and Z. Li, "Illumination-insensitive features for face recognition," *The Visual Computer*, vol. 33, no. 11, pp. 1483–1493, 2017.
- [22] G. Farneback, "Two-frame motion estimation based on polynomial expansion," in *Proceedings of the 13th Scandinavian Conference on Image Analysis*, Halmstad, Sweden, June 2003.
- [23] H. U. Jie, L. I. Shen, S. Albanie et al., "Squeeze-and-excitation networks," in *Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence*, Salt Lake City, UT, USA, June 2019.
- [24] S. Woo, J. Park, L. Joon-Young et al., "CBAM:convolutional block attention module," in *Proceedings of the European Conference on Computer Vision*, ECCV, Munich, Germany, September 2018.
- [25] Z. W. Zhang, J. J. Yan, S. F. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in *Proceedings of the International Conference on Biometrics*, IEEE, New Delhi, India, pp. 26–31, June 2012.
- [26] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of localbinary patterns in face anti-spoofing," in *Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG)*, Hong Kong, China, September 2012.
- [27] Y. A. U. Rehman, L. M. Po, and M. Liu, "Livenet: improving features generalization for face liveness detection using convolution neural networks," *Expert Systems with Applications*, vol. 108, pp. 159–169, 2018.

## Research Article

# Semiparametric Deep Learning Manipulator Inverse Dynamics Modeling Method for Smart City and Industrial Applications

Nan Liu,<sup>1,2</sup> Liangyu Li,<sup>1</sup> Bing Hao,<sup>3</sup> Liusong Yang,<sup>3</sup> Tonghai Hu,<sup>3</sup> Tao Xue,<sup>2</sup> Shoujun Wang ,<sup>2</sup> and Xingmao Shao<sup>2</sup>

<sup>1</sup>School of Mechanical Engineering, Tianjin Polytechnic University, Tianjin 300387, China

<sup>2</sup>National Demonstration Center for Experimental Mechanical and Electrical Engineering Education, Tianjin University of Technology, Tianjin 300384, China

<sup>3</sup>CITIC Heavy Industries Co., Ltd., Luoyang 471003, China

Correspondence should be addressed to Shoujun Wang; wangshoujun@tjut.edu.cn

Received 14 April 2020; Revised 25 May 2020; Accepted 2 June 2020; Published 30 June 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Nan Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In smart cities and factories, robotic applications require high accuracy and security, which depends on precise inverse dynamics modeling. However, the physical modeling methods cannot include the nondeterministic factors of the manipulator, such as flexibility, joint clearance, and friction. In this paper, the Semiparametric Deep Learning (SDL) method is proposed to model robot inverse dynamics. SDL is a type of deep learning framework, designed for optimal inference, combining the Rigid Body Dynamics (RBD) model and Nonparametric Deep Learning (NDL) model. The SDL model takes advantage of the global characteristics of classic RBD and the powerful fitting capabilities of the deep learning approach. Moreover, the parametric and nonparametric parts of the SDL model can be optimized at the same time instead of being optimized separately. The proposed method is validated using experiments, performed on a UR5 robotic platform. The results show that the performance of SDL model is better than that of RBD model and NDL model. SDL can always provide relatively accurate joint torque prediction, even when the RBD or NDL model is not accurate.

## 1. Introduction

Smart cities and factories contain “intelligent” things that can autonomously and collaboratively enhance the quality of living and working conditions, save human lives, and act as a sustainable resource ecosystem. To implement these advanced collaborative technologies, such as drones, robots, artificial intelligence, and Internet of Things, it is required to increase the “intelligence” of smart cities and factories, by improving the connectivity, energy efficiency, and quality of services [1]. There have been many excellent application cases, such as [2–4]. Particularly, intelligent robotic platforms are a technology, increasingly used, in smart cities and factories, where the constantly changing applications scenarios also place higher demands in robot control. Specifically, in motion control systems, there is a time delay in the transmission of feedback information, making smooth

motion impossible to achieve by feedback control alone. Therefore, feedforward control becomes particularly important. In robotics, feedforward control usually refers to model-based control, involving the dynamics of the robotic platform. The accuracy of such dynamical models is critical to the development of control laws that are compliant, energy efficient, and safe [5].

There are two major approaches for modeling robot dynamics: parametric and nonparametric. Parametric approaches rely on parameterized Newtonian physics models of the robot dynamics. Common methods for physics-based dynamics modeling can be found in the literature. These methods require the mechanical parameters of the rigid bodies, composing the robot, to be identified [6–9] and then employed in model-based control and state estimation schemes [10]. The advantage of these models is that they represent a global and unique relationship between the joint

trajectory  $(q, \dot{q}, \ddot{q})$  and the torques  $\tau_{\text{RBD}}$ . This type of inverse dynamics model can be computed efficiently and employed in real time. Thus, a great deal of prior knowledge is acquired, without the need of data. For example, it is well known that robots are subject to gravitational forces, viscous forces, and joint constraints, making it wasteful to have to go through a laborious data-gathering and machine learning process, to discover these well-known constraints. The disadvantage of parametric models is that they are only crude idealizations of the actual system dynamics, such as rigidity of links or a simple analytical form of friction, which may not be accurate in real systems. In the case of traditional industrial robots, these unmodeled dynamics can often be ignored. However, for modern robotic platforms, these omissions and simplifications result in significant control inefficiencies.

Alternatively, the model can be obtained from experimental data, using machine learning techniques, resulting in a nonparametric model. Nonparametric methods, based on algorithms such as Support Vector Regression (SVR) [11–13], Neural Network (NN) [14–16], Local Weighted Projection Regression (LWPR) [17–19], Independent Joint Learning (IJL) [20–22], or Gaussian Processes Regression (GPR) [23–27], can model dynamics by extrapolating the input-output relationship directly from the available data. If a suitable kernel function or learning architecture is selected, then the nonparametric model is a universal approximator which can account for the dynamics factors, not considered by the parametric model. Therefore, nonparametric methods can be more flexible to use and are powerful in capturing higher order nonlinearities, resulting in faster model approximation and higher learning accuracy. When learning inverse dynamics, the nonparametric methods will approximate a function describing the relationship  $q, \dot{q}, \ddot{q} \rightarrow \tau$ , including all nonlinearities encoded by the sampled data.

Nonparametric methods attempt to learn the model from scratch and, thus, do not make use of any knowledge available from analytical robotics. Nevertheless, nonparametric learning methods also exhibit several drawbacks. First, very large amounts of data are necessary for obtaining a sufficiently accurate model and predictions on the entire input space [28]. Second, since nonparametric models rely on local neighborhood training data to make predictions, they do not generalize well to unexplored state regions, where little or no training data are available. Covering the entire state space becomes exponentially harder, as the complexity and number of degrees of freedom in the robot system increase. Thus, if only small and relatively poor data sets are available, nonparametric models will not be able to generalize well for unknown data. Third, it is indeed wasteful to have to go through a laborious data-gathering and machine learning process to discover such well-known prior knowledge as Rigid Body Dynamics.

Thus, it appears quite desirable to combine the benefits of parametric and nonparametric approaches to improve on the aforementioned issues. However, doing so, in an efficient way, is not trivial. A reasonable approach would be to first fit a parametric model and then fit a nonparametric model to

the errors made by the parametric model. Nguyen-Tuong et al. [29] present a learning technique which combines prior knowledge about the physical structure of the mechanical system and learning from available data using Gaussian Process Regression (GPR) [30]. Similar approaches are presented in [20] and [31]. In [32], an incremental semiparametric robot dynamics learning scheme, based on Locally Weighted Projection Regression (LWPR), initialized using a linearized parametric model, is presented [33]. However, this approach uses a fixed parametric model that is not updated, as new data become available. Moreover, LWPR has been shown to underperform with respect to other methods (e.g., [34]). These semiparametric methods, as described above, could not benefit from simultaneous optimization of parametric and nonparametric models. Instead, the nonparametric model is applied, after parametric identification, which may result in a suboptimal model. In addition, as far as it can be known, there is no semiparametric method based on deep learning methods.

Deep learning is a new approach in machine learning, which has been widely applied in smart cities and factories [35]. Deep learning has turned out to be very good at discovering intricate structures in high-dimensional data and is therefore applicable to many domains of science, business, and government. Since it requires very little engineering by hand, it can easily take advantage of increased amount of available computation resources and data [36]. In this work, a method that is based on deep learning and semiparametric approach is presented. The method is formalized in the framework of what is called Semiparametric Deep Learning (SDL), designed for optimal inference using combinations of parametric RBD and Nonparametric Deep Learning models. Key properties of this method are (1) appropriate deep learning frame for a semiparametric approach and (2) features that can be optimized simultaneously for parametric and nonparametric models. The proposed method is validated using experiments performed on a UR5 robot. The article is organized as follows. In Section 2, a complete description of the proposed Semiparametric Deep Learning framework is introduced. Section 3 presents the validation of the proposed method on the UR5 robotic platform. Finally, Section 4 summarizes the content of the presented work.

## 2. Methodology

The parametric modeling method and NDL, as the basis of the SDL method, have been elaborated in previous research publications [37, 38]. Therefore, this section only briefly reviews the above two methods, while it analyzes the SDL modeling method, proposed in this paper.

*2.1. Parametric Robot Dynamics Model.* It is well known that the robot dynamics can be modeled according to the following [39]:

$$\tau(q, \dot{q}, \ddot{q}) = M(q)\ddot{q} + C(q, \dot{q}) + G(q) + \varepsilon(q, \dot{q}, \ddot{q}), \quad (1)$$

in which  $q, \dot{q}, \ddot{q} \in \mathbb{R}^{m \times 1}$  are joint positions, velocities, and accelerations of the robot, respectively,  $\tau \in \mathbb{R}^{m \times 1}$  denotes the joint torques,  $M(q)$  is the generalized inertia matrix of the robot,  $C(q, \dot{q})$  are the Coriolis and centripetal forces, and  $G(q)$  is gravity. As shown in equation (1), the robot dynamics equation contains Rigid Body Dynamics (RBD) model:

$$\tau_{\text{RBD}} = M(q)\ddot{q} + C(q, \dot{q}) + G(q). \quad (2)$$

The model errors  $\varepsilon(q, \dot{q}, \ddot{q})$  are caused by unmodeled dynamics (e.g., hydraulic tubes, actuator dynamics, and flexibility and dynamics of the cable drives), ideal-joint assumptions (e.g., no friction and clearance), and inaccuracies in the RBD model parameters. The RBD model of a manipulator is well known to be linear regarding the parameters  $\beta$  [39], i.e.,

$$\tau_{\text{RBD}} = \varphi(q, \dot{q}, \ddot{q})\beta, \quad (3)$$

in which  $\varphi$  is a matrix containing nonlinear functions of joint angles, velocities, and accelerations, often called basis functions. Modeling the robot dynamics, using the RBD model in equation (3), requires the identification of the dynamics parameters  $\beta$ . For the 6 Degree-of-Freedom (DoF) UR5 robot, for example, 60 dynamics parameters are to be identified (for each DoF, there are 10 parameters that could ideally be obtained directly from the CAD data).

## 2.2. Nonparametric Deep Learning Model

**2.2.1. NDL Formulation.** The inverse dynamic model, in model-based control, is described as the mapping from joint positions, velocities, and accelerations to torques, as shown in equation (1). The aim of nonparametric learning model is to predict the torque value of the  $i^{\text{th}}$  joint,  $\tau_i \in \mathbb{R}$ , as the response of the query point at  $(q, \dot{q}, \ddot{q}) \in \mathbb{R}^{3m \times 1}$ , by using the given  $n$  training data,  $\{q[j], \dot{q}[j], \ddot{q}[j], \tau_i[j]\}_{j=1}^n$ , in which  $q[j], \dot{q}[j], \ddot{q}[j] \in \mathbb{R}^{m \times 1}$  and  $\tau_i[j] \in \mathbb{R}$ . Since the problem can be considered as a supervised learning problem, any supervised learning technique can be employed for the learning process, as shown in

$$\tau(q, \dot{q}, \ddot{q}) \sim \mathcal{DL}(q, \dot{q}, \ddot{q}). \quad (4)$$

Deep learning methods are representation-learning methods with multiple levels of representation, obtained by composing simple but nonlinear modules, each transforming the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level. With the composition of enough such transformations, very complex robotic dynamics functions can be learned [36, 40]. Therefore, the application cases of deep learning are too numerous, such as [41–45].

The sequential nature of manipulator inverse dynamics suggests that, to predict the joint torque, it is important to model the relationship among sequential data points [39]. RNNs, a type of deep learning network, can be seen as very deep feedforward network, where all the layers share the same weights. Although their main purpose is to determine long-term dependencies, theoretical and empirical evidence

shows that it is difficult to learn to store information for very long time [46]. LSTM networks have subsequently proved to be more effective than conventional RNNs, especially when there are several layers at each time step [48], enabling an entire speech recognition system that goes all the way from acoustics to the sequence of characters in the transcription. LSTM networks or related forms of gated units are also currently used for the encoder and decoder networks, performing very well in machine translation [47–49]. Moreover, studies presented in [38] and [50] confirmed the validity of applying LSTM to the prediction of manipulator inverse dynamics. Therefore, in this section, the LSTM network is proposed as the nonparametric learning technique for modeling the inverse dynamics of manipulator.

**2.2.2. NDL Model Architecture.** In this paper, the proposed architecture of the Nonparametric Deep Learning network has one input layer, one LSTM layer, one full-connected layer, one dropout layer, and one output layer, as shown in Figure 1. Only 1 LSTM layer is used here, because it has been verified in our previous studies that, with the same number of neurons, the fewer the layers, the better the prediction performance [38].

The input layer has 18 neurons (manipulator's 6 joint positions, 6 velocities, and 6 accelerations, as shown in Figure 2).

The state activation functions of LSTM cells are set to "tanh," while the gate activation functions are set to "sigmoid." The input weights are initialized according to the Glorot initializer. The forget gate bias is initially set to 1 and the remaining biases are set to 0. The training algorithm adopts back-propagation through time.

## 2.3. Proposed Semiparametric Deep Learning (SDL) Model.

In this section, the proposed semiparametric model, based on deep learning and RBD, is described in detail. The method is formalized in the framework of the so-called Semiparametric Deep Learning (SDL), which can be used to predict the joint torque of a robotic arm more accurately. First, formulation of SDL is introduced in Section 2.3.1, while the specific model architecture of SDL is described in Section 2.3.2.

**2.3.1. SDL Formulation.** Using the Nonparametric Deep Learning (DL) framework, the robot dynamics can be modeled by  $\tau \sim \mathcal{DL}(q, \dot{q}, \ddot{q})$ , in which  $q, \dot{q}, \ddot{q}$  is the input and  $\tau$  is the output of the Deep Learning Model. Consequently, the DL model does not make use of any prior knowledge, which allows reproducing arbitrary functions. One way to include the RBD model, as shown in equation (5), is to set the  $\tau_{\text{RBD}}$  as input to the DL model. This approach is equivalent to a semiparametric model:

$$\begin{aligned} \tau(q, \dot{q}, \ddot{q}) &\sim \mathcal{SDL}(q, \dot{q}, \ddot{q}) \sim \mathcal{DL}(\tau_{\text{RBD}}, q, \dot{q}, \ddot{q}) \\ &\sim \mathcal{DL}(\varphi(q, \dot{q}, \ddot{q})\beta, q, \dot{q}, \ddot{q}). \end{aligned} \quad (5)$$

The resulting dynamics model, as described in equation (5), is a semiparametric frame which consists of a parametric



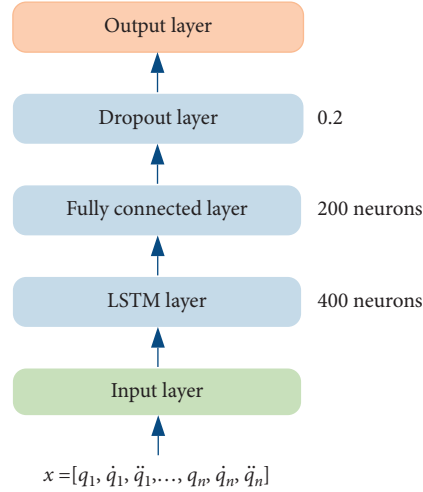


FIGURE 1: Proposed deep learning architecture.

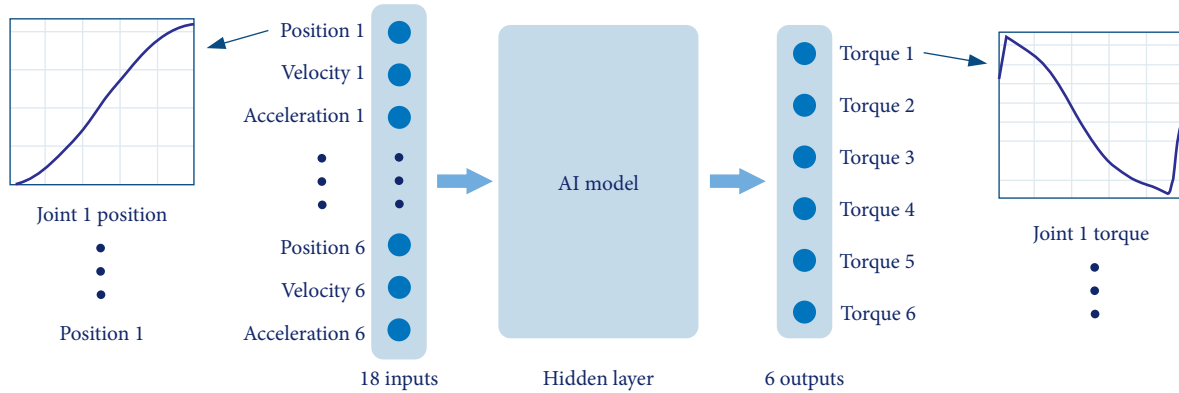


FIGURE 2: Inverse dynamics learning architecture of manipulator with 21 input data sets.

part, i.e., the RBD model. When comparing equation (5) to the robot dynamics in equation (1), it becomes evident that the main purpose of the nonparametric term is absorbing the unmodeled dynamics  $\varepsilon(q, \dot{q}, \ddot{q})$ . In order to approximate the unmodeled dynamics with an appropriate DL, a model can be used, such as DNN or LSTM. Key properties of this semiparametric learning method are (1) employment of appropriate deep learning frame as nonparametric part and (2) features that can be optimized simultaneously for parametric and nonparametric model; i.e., the value of  $\tau_{\text{RBD}}$  will also be updated by weight changes as the learning process evolves.

If the RBD model perfectly describes the robot dynamics, the error  $\varepsilon(q, \dot{q}, \ddot{q})$  in equation (1) will disappear and the prediction will depend only on the RBD part; i.e., it is very easy to train deep learning networks. Equation (5) also shows that if the query point is far away from the training data, the resulting torque prediction will mainly depend on the RBD part. This property is important, as the complete state space can never be completely included using finite (and possibly small) training data sets. If the robot moves to the regions of the state space, not considered by the sampled data (i.e., the learned nonparametric models may not

generalize well in these state space regions), the torque prediction will rely on the parametric RBD part.

**2.3.2. SDL Model Architecture.** The effectiveness of the NDL model has been verified in previous work presented in [40]. The SDL model proposed in this article adds an RBD term to the NDL architecture. Its specific model architecture is shown in Figure 3. The input of the SDL architecture is still  $[q, \dot{q}, \ddot{q}]$ . However, unlike NDL, the input passes through the RBD term, forming the new input vector  $[\tau_{\text{RBD}}, q, \dot{q}, \ddot{q}]$ . Next, the new input vector enters the LSTM hidden layer. The advantage of this architecture is that the parametric and nonparametric parts can be optimized simultaneously; i.e., the weight of  $\tau_{\text{RBD}}$  in the feature vector will also be updated during network training. The grid search method is used to optimize the hyperparameters of the SDL model.

### 3. Evaluation

The proposed SDL method will be verified on a collaborative robot UR5, while the torque prediction results will be compared to the ones provided by NDL methods. The prediction performance for training and for generating

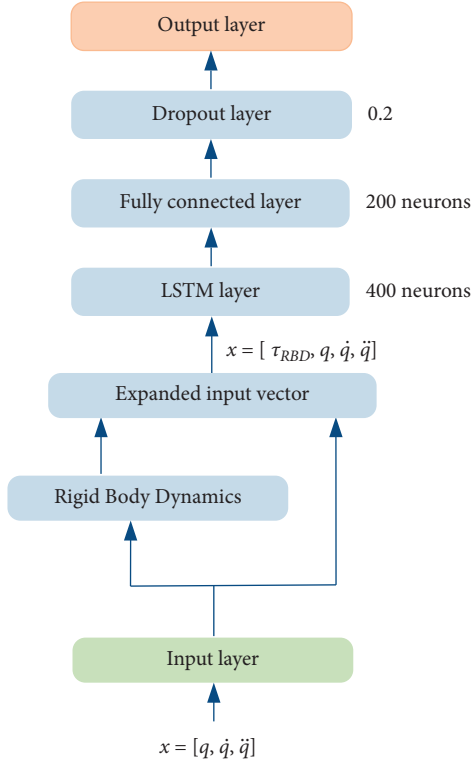


FIGURE 3: Proposed deep learning architecture.

predictions in rhythmic motor tasks is evaluated. The joint angles, the joint velocities, the joint accelerations, and the joint torques are recorded using a GUI (Graphical User Interface), i.e., PolyScope, making it easy to program the robot to move the tool along a desired trajectory path.

**3.1. Experimental Setup.** UR5 is a 6-DoF collaboration robot with extruded aluminum tubes and joints. It has six rotary joints, and its structure is shown in Figure 4. The UR5 robot has a joint rotation range of  $[-2\pi, 2\pi]$  (rads) and a joint acceleration range of  $[0, \pi]$  (rads/s<sup>2</sup>). The UR5 robot is very popular in the robot research field.

According to the robot rotation angle and installation restrictions, the robot workspace is selected to be a hemisphere with an approximate radius of 850 mm above the installation plane. The range of joints motion is shown in Table 1. In order to best approximate the actual working situation of the robot, within the selected robot workspace, 1000 points are randomly selected. According to the actual use requirements of the robot, the joint running speed range is  $[0.8-2]$  rads/s, while the acceleration range is  $[1-1.8]$  rads/s.

The robot is ordered to run in a rhythmic way, according to the set trajectory. The joint position, speed, and servo motor current data, along the robot trajectory, is delivered from the robot controller at a frequency of 100 Hz. Since the UR5 robot is not equipped with a torque sensor, the measured torque is obtained indirectly through the motor current, at each joint. The relationship between torque and current is as follows [51]:

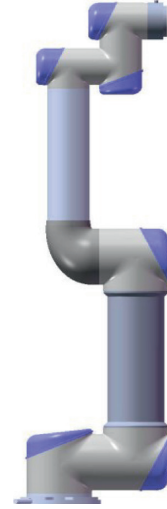


FIGURE 4: UR5 robot structure.

$$\tau_i = N_i k_i I_i, \quad (6)$$

in which  $N_i$  is the gear ratio,  $N_i = 101, i = 1, \dots, 6$ ;  $k_i$  motor constant,  $k_i = 0.125 \text{ Nm/A}, i = 1, \dots, 3$ ,  $k_i = 0.0922 \text{ Nm/A}, i = 4, \dots, 6$ ; and  $I_i$  is the motor current (A).

During the actual operation, the robot is affected by noise, making the sampling data fluctuate. In that case, there will be large fluctuations and ripples in the actual measured current, which seriously affects the accuracy of the torque prediction. The average data method can increase the signal-to-noise ratio of the data [52], reduce the influence of noise, and improve the prediction accuracy. The average joint position  $\bar{q}$  can be expressed as follows:

$$\bar{q}(k) = \frac{1}{M} \sum_{m=1}^M q_m(k), \quad (7)$$

in which  $M$  is the number of running trajectories,  $q_m(k)$  is the  $k$ -th sampling point of one running trajectory, and  $\bar{q}(k)$  is the position after  $M$  times of averaging. The same method is used to deal with speed and current. The zero-phase low-pass Butterworth filter (forward and reverse IIR Butterworth filters) with a cutoff low-pass frequency of 1 Hz is used to process the averaged position and velocity  $(\bar{q}, \dot{\bar{q}})$ . Acceleration is obtained by the central difference method [53]. Data processing is important for the accuracy of parameter identification, making the above processing very appropriate, as it avoids large deviations in identification.

A total of 1,000 groups (a total of 108,008) of valid samples were obtained. According to the ratio of 80% to 20%, they were divided into training and testing sets, serving in  $K$ -fold cross-validation. In addition, the test data is ensured to be sufficiently different from the training data, highlighting the generalization ability of the learned models. The above sample set is used to train and test the NDL model and SDL model, respectively, while it also analyzes and compares the prediction performance.

TABLE 1: Rotation angle range of each joint in the selected robot workspace.

| Joint #               | Joint 1     | Joint 2   | Joint 3     | Joint 4   | Joint 5     | Joint 6     |
|-----------------------|-------------|-----------|-------------|-----------|-------------|-------------|
| Range of motion (deg) | [-180, 180] | [-180, 0] | [-130, 130] | [-180, 0] | [-180, 180] | [-180, 180] |

TABLE 2: Prediction results of RBD, NDL, and SDL models.

|               |     | Axis 1 | Axis 2 | Axis 3 | Axis 4 | Axis 5 | Axis 6 | All axes |
|---------------|-----|--------|--------|--------|--------|--------|--------|----------|
| Cross 1       | RBD | 2.2899 | 2.3417 | 2.3683 | 0.6523 | 0.5889 | 0.5290 | 1.7024   |
|               | NDL | 2.6107 | 2.7715 | 1.8227 | 0.3418 | 0.4487 | 0.4577 | 1.7486   |
|               | SDL | 1.8310 | 1.9460 | 1.3504 | 0.3450 | 0.4220 | 0.4524 | 1.2560   |
| Cross 2       | RBD | 2.3316 | 2.3162 | 2.2893 | 0.6642 | 0.6498 | 0.5525 | 1.6937   |
|               | NDL | 2.3342 | 2.5439 | 2.5428 | 0.3864 | 0.5162 | 0.4722 | 1.7807   |
|               | SDL | 1.6721 | 1.8492 | 1.4528 | 0.3435 | 0.4633 | 0.4654 | 1.2162   |
| Cross 3       | RBD | 2.4861 | 2.6723 | 2.2126 | 0.6417 | 0.5924 | 0.3582 | 1.7846   |
|               | NDL | 2.2477 | 2.4820 | 1.1689 | 0.4081 | 0.4487 | 0.3977 | 1.4779   |
|               | SDL | 1.4031 | 1.5533 | 0.9656 | 0.3294 | 0.4022 | 0.3428 | 0.9748   |
| Cross 4       | RBD | 2.7325 | 2.7323 | 2.5119 | 0.6625 | 0.6386 | 0.3708 | 1.9246   |
|               | NDL | 2.3239 | 2.1714 | 1.6374 | 0.3745 | 0.4245 | 0.4280 | 1.4889   |
|               | SDL | 1.4009 | 1.6816 | 1.1613 | 0.3435 | 0.4399 | 0.3696 | 1.0478   |
| Cross 5       | RBD | 2.8011 | 2.7946 | 2.2741 | 0.6326 | 0.5795 | 0.3620 | 1.9015   |
|               | NDL | 1.6768 | 2.2989 | 1.6690 | 0.4346 | 0.3893 | 0.3393 | 1.3746   |
|               | SDL | 1.5326 | 1.6847 | 1.0004 | 0.3385 | 0.3775 | 0.3058 | 1.0439   |
| Cross average | RBD | 2.5282 | 2.5714 | 2.3313 | 0.6506 | 0.6098 | 0.4345 | 1.8014   |
|               | NDL | 2.2387 | 2.4535 | 1.7682 | 0.3891 | 0.4455 | 0.4190 | 1.5741   |
|               | SDL | 1.5679 | 1.7430 | 1.1861 | 0.3400 | 0.4210 | 0.3872 | 1.1077   |

All input and output data are normalized to match the consistency of the learning model. After the prediction of inverse dynamics, the real value is restored. The normalized equation is as follows:

$$x_n = \frac{x_r - x_{\min}}{x_{\max} - x_{\min}}, \quad (8)$$

in which  $x_n$  represents the normalized value;  $x_r$  denotes the real value; and  $x_{\min}$  and  $x_{\max}$  are the minimum and maximum real values, respectively.

The performance of manipulator inverse dynamics predictions is evaluated using Root Mean Square Error (RMSE), which is defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (p_i - y_i)^2}, \quad (9)$$

in which  $p_i$  and  $y_i$  represent the  $i$ -th predicted value and real value, respectively, and  $N$  is the total number of the data sets.

## 4. Results

The training and prediction in this paper were performed with MATLAB 2019a, using an ordinary personal computer. Computer hardware has a high influence on training time. In this work, the models are trained on a CPU with a clock speed of 2.7 GHz. The structure and hyperparameters of the NDL and SDL models were initially set according to previous work, while the final settings were determined based on the fivefold cross-validation method.

The prediction results of the RBD, NDL, and SDL models are listed in Table 2. The data in the table shows the RMSE of the RBD, NDL, and SDL models, on different robot axes and under different cross situations. Based on Table 2, the following conclusions apply:

- (1) The results of “cross average” (Figure 5) indicate that the prediction accuracy of SDL is generally more accurate than the one of NDL and NDL is generally more accurate than RBD. Also, the prediction accuracy of the SDL model for the first 3 axes of the robot is significantly improved.
- (2) The results of “all axes” (Figure 6) indicate that the prediction accuracy of the NDL model is sometimes better than the one of the RBD model (cross 3, 4, 5), while sometimes it is worse (cross 1, 2). However, the NDL model is better than the RBD model (mean) in general, while the prediction accuracy of the SDL model is always better than in the case of the other two models.
- (3) All the data in Table 2 shows that the semiparametric models are able to combine the strengths of both models, i.e., the parametric RBD model and the Nonparametric Deep Learning model. The prediction accuracy of the SDL model is always better than that of the RBD and NDL models.

RMSE is not sufficient to fully represent the performance of torque prediction, because the range of torque variation for each joint of the robot varies greatly. Therefore, the ratio of cross average RMSE to the range of measured torque values was used to further analyze the predictive

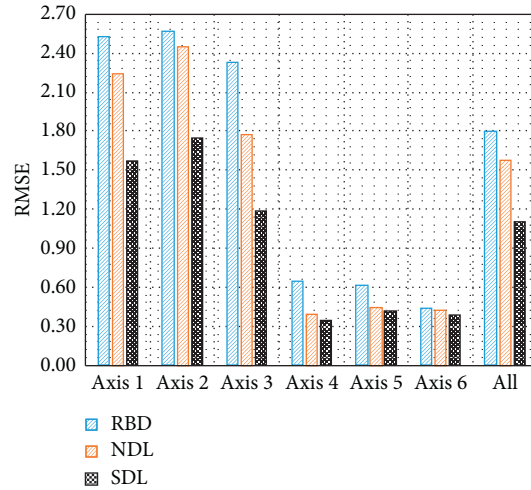


FIGURE 5: Cross average prediction results.

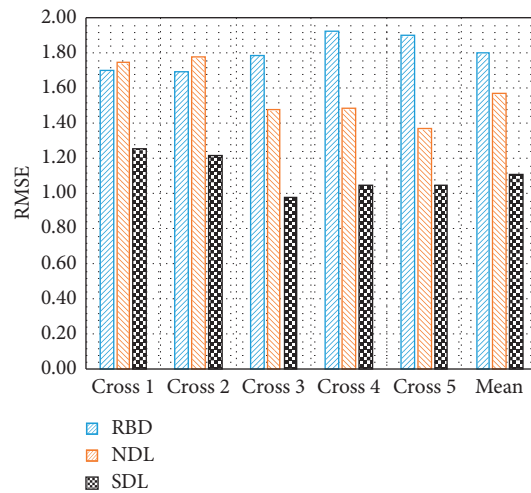


FIGURE 6: All axes prediction results.

TABLE 3: Comparative analysis of prediction performance of different joint forces.

|                          |     | Axis 1  | Axis 2  | Axis 3  | Axis 4 | Axis 5 | Axis 6 |
|--------------------------|-----|---------|---------|---------|--------|--------|--------|
| Cross average RMSE       | RBD | 2.5282  | 2.5714  | 2.3313  | 0.6506 | 0.6098 | 0.4345 |
|                          | NDL | 2.2387  | 2.4535  | 1.7682  | 0.3891 | 0.4455 | 0.4190 |
|                          | SDL | 1.5679  | 1.7430  | 1.1861  | 0.3400 | 0.4210 | 0.3872 |
| Range of measured torque |     | 42.1816 | 98.1207 | 49.3103 | 5.7766 | 4.4944 | 6.6976 |
| Ratio                    | RBD | 0.0599  | 0.0262  | 0.0473  | 0.1126 | 0.1357 | 0.0649 |
|                          | NDL | 0.0531  | 0.0250  | 0.0359  | 0.0674 | 0.0991 | 0.0626 |
|                          | SDL | 0.0372  | 0.0178  | 0.0241  | 0.0589 | 0.0937 | 0.0578 |

performance of different joints. As shown in Table 3, the torque prediction performance of the first three (elbow) joints is better than that of the last three (wrist) joints.

Part of the measured values of each joint torque, RBD predicted value, NDL predicted value, and SDL predicted value are all plotted in Figure 7. Figures 7(a)–7(f) are the predicted moments of axis 1 to axis 6, respectively; Figures 7(g)–7(l) are the predicted errors of moments of axis

1 to axis 6, respectively; the purple curve represents the measured torque value, the blue represents the calculated torque value using RBD method, the red represents the DL predicted torque value, and the yellow represents the SDL predicted torque value. Figure 7 shows that the SDL model combines the advantages of the RBD model and the NDL model. For example, as shown by the red dashed box in Figure 7, when the torque prediction error of the RBD is

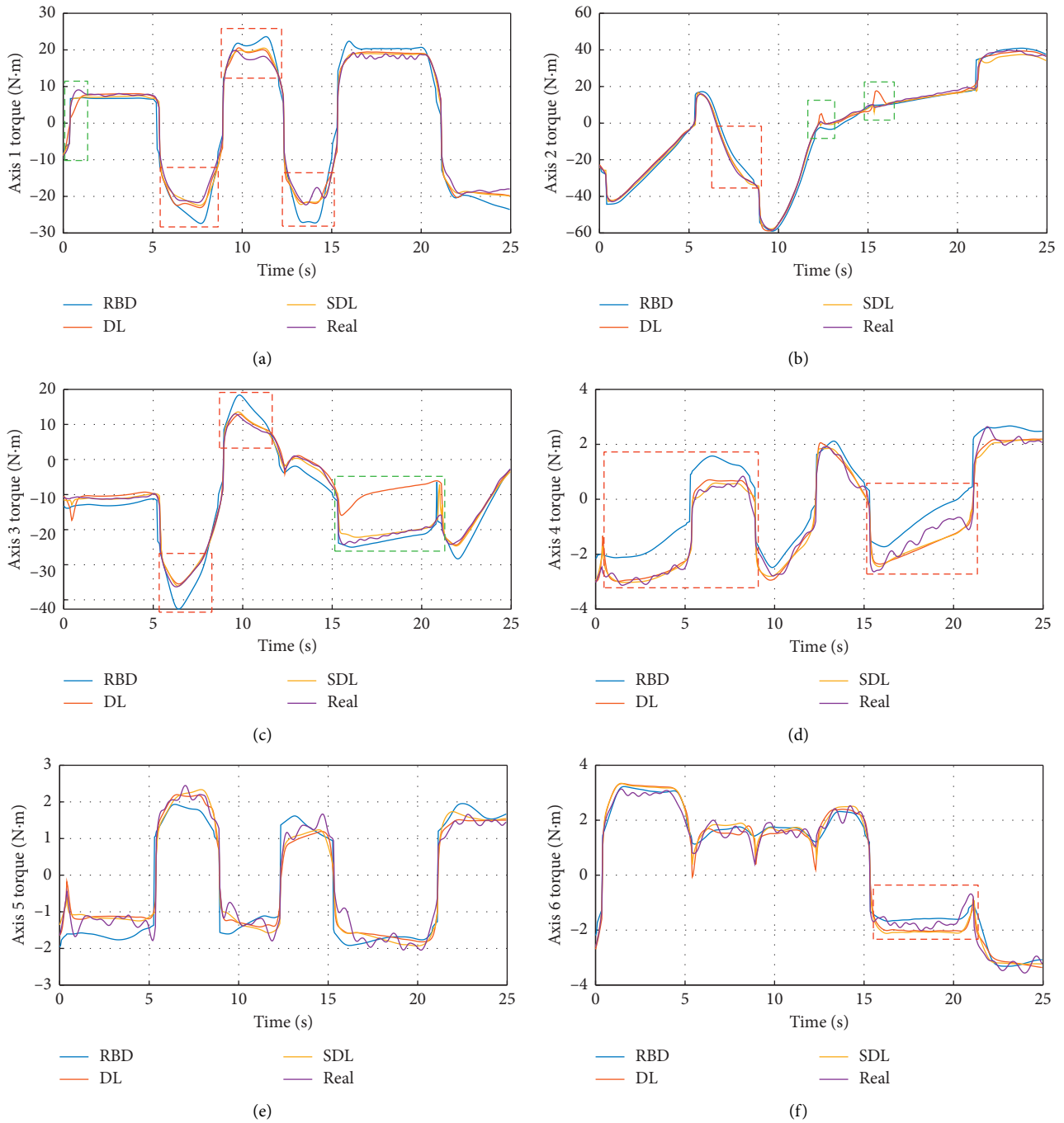


FIGURE 7: Continued.

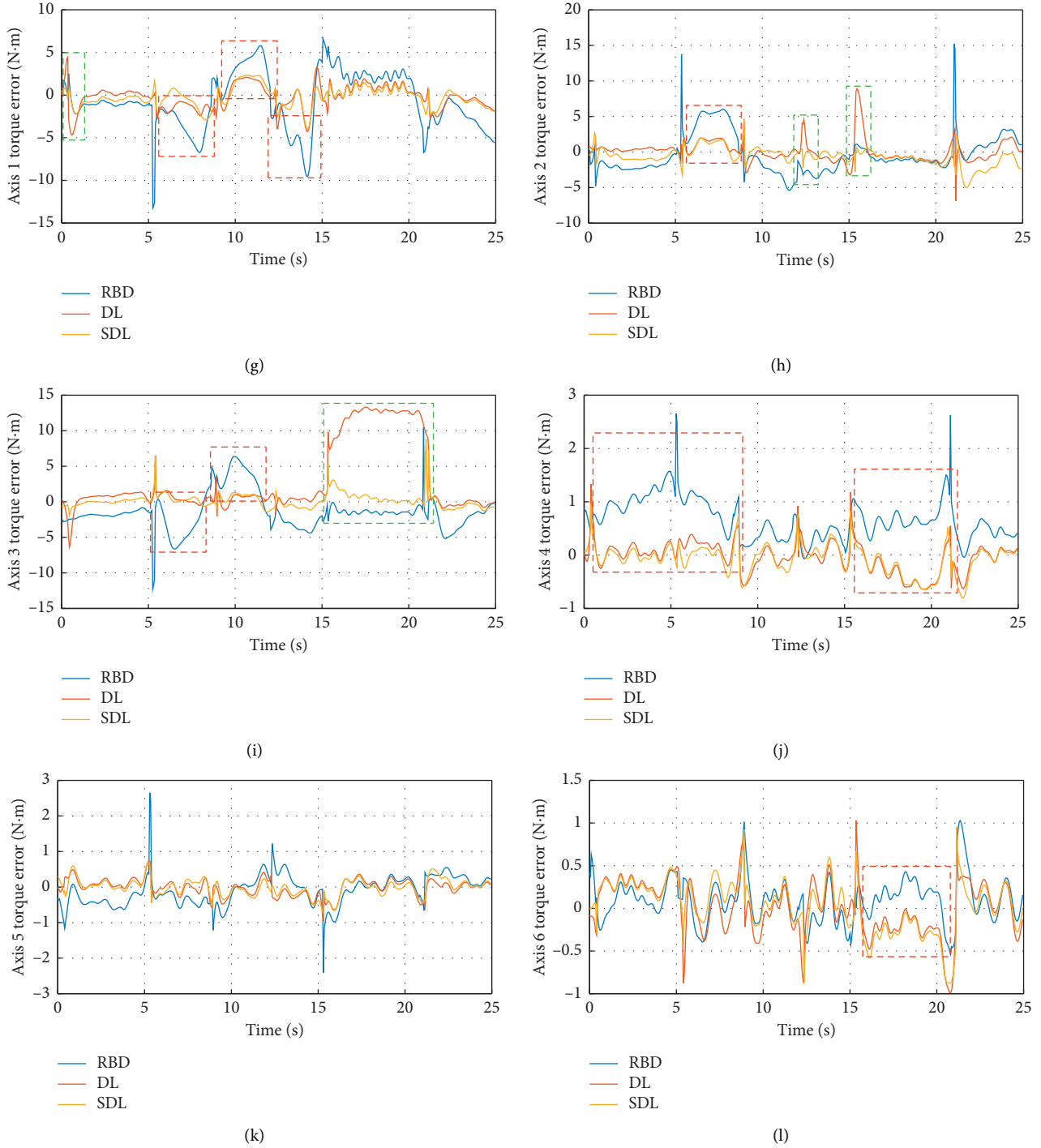


FIGURE 7: Joint torque prediction comparison curves. (a) Axis 1 torque. (g) Axis 1 torque error. (b) Axis 2 torque. (h) Axis 2 torque error. (c) Axis 3 torque. (i) Axis 3 torque error. (d) Axis 4 torque. (j) Axis 4 torque error. (e) Axis 5 torque. (k) Axis 5 torque error. (f) Axis 6 torque. (l) Axis 6 torque error.

large, the nonparametric part works, which greatly improves the prediction accuracy of the SDL model. As another example, shown by the green dotted box in Figure 7, when the learned nonparametric models do not generalize well to the state space regions, the torque prediction will rely on the parametric RBD part.

## 5. Conclusion

In this work, Semiparametric Deep Learning (SDL) method is proposed to model robot inverse dynamics, for smart city and industrial applications. The SDL model takes advantage of the global characteristics of classic RBD and the powerful

fitting capabilities of deep learning methods. Moreover, SDL model can be optimized simultaneously for the parametric and nonparametric model, instead of separate optimizations. The results on the UR5 robot show that the SDL models provide higher accuracy and better generalization, compared to RBD and NDL. The essence of the SDL model is to fully utilize the a priori information encoded in the parameterized model, overcome the limitations of the NDL method, and always show good learning performance. As far as future work is concerned, the flexible factors in dynamics will be considered as addition to the SDL model, in order to improve the accuracy prediction performance.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the National Key Research and Development Program of China (no. 2017YFB1302100).

## References

- [1] J. Jiang, F. Lin, J. Fan et al., "A destination prediction network based on spatiotemporal data for bike-sharing," *Complexity*, vol. 2019, Article ID 7643905, 14 pages, 2019.
- [2] G. Li, L. Zhang, Y. Sun, and J. Kong, "Towards the sEMG hand: internet of things sensors and haptic feedback application," *Multimedia Tools and Applications*, vol. 78, no. 21, pp. 29765–29782, 2019.
- [3] Y. He, G. Li, Y. Liao et al., "Gesture recognition based on an improved local sparse representation classification algorithm," *Cluster Computing*, vol. 22, no. S5, pp. 10935–10946, 2019.
- [4] J. Qi, G. Jiang, G. Li, Y. Sun, and B. Tao, "Intelligent human-computer interaction based on surface EMG gesture recognition," *IEEE Access*, vol. 7, pp. 61378–61387, 2019.
- [5] B. Siciliano and O. Khatib, *Dynamics*, Springer Handbook of Robotics, Springer, Berlin, Germany, 2008.
- [6] K. Yamane, "Practical Kinematic and Dynamic Calibration Methods For Force-Controlled Humanoid Robots," in *Proceedings of 11th IEEE-RAS International Conference on Humanoid Robots*, ICMLBled, Slovenia, pp. 269–275, 2011.
- [7] S. Traversaro, A. D. Prete, R. Muradore, L. Natale, and F. Nori, "Inertial Parameter Identification Including Friction and Motor Dynamics," in *Proceedings of 13th IEEE-RAS International Conference on Humanoid Robots*, pp. 68–73, Atlanta, GA, USA, 2013.
- [8] Y. Ogawa, G. Venture, and C. Ott, "Dynamic Parameters Identification Of A Humanoid Robot Using Joint Torque Sensors and/or Contact Forces," in *Proceedings of 14th IEEE-RAS International Conference on Humanoid Robots*, pp. 457–462, Atlanta, GA, USA, 2014.
- [9] J. Hollerbach, W. Khalil, and M. Gautier, "Model Identification," *Springer Handbook of Robotics*, Springer, Berlin, Germany, 2008.
- [10] R. Camoriano, S. Traversaro, L. Rosasco et al., "Incremental semiparametric inverse dynamics learning," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, Paris, France, pp. 544–550, 2016.
- [11] Y.-P. Zhao, B. Li, Y.-B. Li, and K.-K. Wang, "Householder transformation based sparse least squares support vector regression," *Neurocomputing*, vol. 161, pp. 243–253, 2015.
- [12] D. Nguyen-Tuong, B. Scholkopf, and J. Peters, "Sparse online model learning for robot control with support vector regression," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, Madrid, Spain, pp. 3121–3126, 2009.
- [13] Y. Choi, S.-Y. Cheong, and N. Schweighofer, "Local online support vector regression for learning control," in *Proceedings of the 2007 IEEE International Symposium on Computational Intelligence in Robotics and Automation*, IEEE, Jacksonville, FL, USA, pp. 13–18, June 2007.
- [14] H. Mori, Y. Ohama, N. Fukumura, and Y. Uno, "Learning of real robot's inverse dynamics by a forward-propagation learning rule," *Electrical Engineering in Japan*, vol. 161, no. 4, pp. 38–48, 2007.
- [15] N. Ishibashi and Y. Maeda, "Learning of inverse-dynamics for SCARA robot," in *Proceedings of the SICE Annual Conference*, pp. 1300–1303, Takamatsu, Japan, September 2011.
- [16] O. Ken and M. Yutaka, "Learning of inverse-dynamics and inverse-kinematics for two-link SCARA robot using neural networks," in *Proceedings of the SICE Annual Conference*, pp. 1031–1034, Takamatsu, Japan, September 2011.
- [17] S. Vijayakumar, A. D'Souza, and S. Schaal, "Incremental online learning in high dimensions," *Neural Computation*, vol. 17, no. 12, pp. 2602–2634, 2005.
- [18] S. Vijayakumar and S. Schaal, "Fast and efficient incremental learning for high-dimensional movement systems," in *Proceedings of the IEEE International Conference on Robotics and Automation*, vol. 2, IEEE, Montreal, Canada, pp. 1894–1899, 2000.
- [19] J. S. d. l. Cruz, *Learning Inverse Dynamics for Robot Manipulator Control*, University of Waterloo, Waterloo, Canada, 2011.
- [20] T. T. Um, M. S. Park, and J. Park, "Independent Joint Learning: a novel task-to-task transfer learning scheme for robot models," 2014," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, Paris, France, pp. 5679–5684, 2014.
- [21] K. Caluwaerts and J. J. Steil, "Independent joint learning in practice: local error estimates to improve inverse dynamics control," in *Proceedings of the 2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*, IEEE, Seoul, South Korea, pp. 643–650, November 2015.
- [22] Z. Shareef, P. Mohammadi, and J. Steil, "Improving the inverse dynamics model of the KUKA LWR IV+ using independent joint learning," Z. Shareef received funding from the German federal ministry of education and research (BMBF) within the leading-edge cluster competition. P. Mohammadi received funding from the European community's horizon 2020 robotics program ICT-23-2014 under grant agreement 644727—CoglMon," *IFAC-PapersOnLine*, vol. 49, no. 21, pp. 507–512, 2016.
- [23] K. Chai, *Multi-task Learning with Gaussian Processes*, Institute for Adaptive and Neural Computation, School of Informatics, University of Edinburgh, Edinburgh, Scotland, 2010.
- [24] M. Deisenroth, J. Peters, and C. Rasmussen, "Approximate dynamic programming with Gaussian processes," in

- Proceedings of the American Control Conference*, pp. 4480–4485, Denver, Colorado, 2008.
- [25] M. Deisenroth and C. R. Pilco, “A model-based and data-efficient approach to policy search,” in *Proceedings of the Twenty Eighth International Conference on Machine Learning*, ICML, Bellevue, WA, USA, 2011.
- [26] D. Nguyen-Tuong and J. Peters, “Local Gaussian process regression for real-time model-based robot control,” in *Proceedings of the International Conference on Intelligent Robots and Systems*, pp. 380–385, Macau, China, 2008.
- [27] A. Rottmann and W. Burgard, “Learning non-stationary system dynamics online using Gaussian processes,” *Pattern Recognition of Lecture Notes in Computer Science*, vol. 6376, pp. 192–201, Springer, Berlin, Germany, 2010.
- [28] D. Nguyen-Tuong, J. Peters, and M. Seeger, “Computed torque control with nonparametric regression models,” in *Proceedings of the 2008 American Control Conference*, ACC, Seattle, DC, USA, 2008.
- [29] D. Nguyen-Tuong and J. Peters, “Using model knowledge for learning inverse dynamics,” in *Proceedings of the 2010 IEEE International Conference on Robotics and Automation*, pp. 2677–2682, Anchorage, AK, USA, May 2010.
- [30] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, MS, USA, 2006.
- [31] T. Wu and J. Movellan, *Semi-Parametric Gaussian Process for Robot System Identification*, IROS, Bengaluru, India, 2012.
- [32] J. Sun de la Cruz, D. Kulic, W. Owen, E. Calisgan, and E. Croft, “On-line dynamic model learning for manipulator control,” *IFAC Robot Control*, vol. 10, no. 1, pp. 869–874, 2012.
- [33] S. Vijayakumar and S. Schaal, “Locally weighted projection regression,” in *Incremental Real Time Learning in High Dimensional Space*. ICML, P. Langley, Ed., pp. 1079–1086, Morgan Kaufmann, Burlington, MS, USA, 2000.
- [34] A. Gijsberts and G. Metta, “Incremental Learning Of Robot Dynamics Using Random Features,” in *Proceedings of the 2011 IEEE International Conference on Robotics and Automation*, pp. 951–956, Shanghai, China, May 2011.
- [35] L. Zhang, C. P. Lim, and J. Han, “Complex deep learning and evolutionary computing models in computer vision,” *Complexity*, vol. 2019, Article ID 1671340, 2 pages, 2019.
- [36] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [37] N. Liu, L. Li, B. Hao et al., “Inverse dynamic modeling and simulation of multiple-degree-of-freedom heavy-duty hydraulic manipulator,” *International Journal of Mechatronics and Applied Mechanics*, vol. 2, no. 6, pp. 87–96, 2019.
- [38] N. Liu, L. Li, B. Hao et al., “Modeling and simulation of robot inverse dynamics using LSTM-based deep learning algorithm for smart cities and factories,” *IEEE Access*, vol. 7, pp. 173989–173998, 2019.
- [39] M. W. Spong, S. Hutchinson, and M. Vidyasagar, *Robot Dynamics and Control*, John Wiley & Sons, New York, NJ, USA, 2006.
- [40] A. Hernandez-Blanco, B. Herrera-Flores, D. Tomas et al., “A systematic review of deep learning approaches to educational data mining,” *Complexity*, vol. 2019, Article ID 1306039, 22 pages, 2019.
- [41] G. Li, J. Li, Z. Ju, Y. Sun, and J. Kong, “A novel feature extraction method for machine learning based on surface electromyography from healthy brain,” *Neural Computing and Applications*, vol. 31, no. 12, pp. 9013–9022, 2019.
- [42] W. Cheng, Y. Sun, G. Li, G. Jiang, and H. Liu, “Jointly network: a network based on CNN and RBM for gesture recognition,” *Neural Computing and Applications*, vol. 31, no. S1, pp. 309–323, 2019.
- [43] G. Li, D. Jiang, Y. Zhou, G. Jiang, J. Kong, and G. Manogaran, “Human lesion detection method based on image information and brain signal,” *IEEE Access*, vol. 7, pp. 11533–11542, 2019.
- [44] D. Jiang, G. Li, Y. Sun, J. Kong, and B. Tao, “Gesture recognition based on skeletonization algorithm and CNN with ASL database,” *Multimedia Tools and Applications*, vol. 78, no. 21, pp. 29953–29970, 2019.
- [45] F. P. An, “Pedestrian re-recognition algorithm based on optimization deep learning-sequence memory model,” *Complexity*, 2019.
- [46] Y. Bengio, P. Simard, and P. Frasconi, “Learning long-term dependencies with gradient descent is difficult,” *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994.
- [47] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to sequence learning with neural networks,” *Advances in Neural Information Processing Systems*, pp. 3104–3112, 2014.
- [48] K. Cho, B. Van Merriënboer, C. Gulcehre et al., “Learning phrase representations using RNN encoder-decoder for statistical machine translation,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1724–1734, Doha, Qatar, October 2014.
- [49] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” in *Proceedings of the 3rd International Conference on Learning Representations*, ICLR, San Diego, CA, USA, 2015.
- [50] E. Rueckert, M. Nakatenus, S. Tosatto et al., “Learning Inverse Dynamics Models in on Time with LSTM Networks,” 2017,” in *Proceedings of the IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*, pp. 811–816, Santa Monica, CA, USA, November 2017.
- [51] N. Kovincic, A. Müller, H. Gattringer et al., “Dynamic parameter identification of the universal robots UR5,” in *Proceedings of the ARW & OAGM Workshop*, pp. 44–53, Steyr, Austrian, May 2019.
- [52] S. Jiang, M. Jiang, Y. Cao et al., “A typical dynamic parameter identification method of 6-degree-of-freedom industrial robot,” *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 231, no. 9, pp. 740–752, 2017.
- [53] J. Jin and N. Gans, “Parameter identification for industrial robots with a fast and robust trajectory design approach,” *Robotics and Computer-Integrated Manufacturing*, vol. 31, pp. 21–29, 2015.



## Research Article

# An Anonymous Authentication Scheme in VANETs of Smart City Based on Certificateless Group Signature

Yuanpan Zheng <sup>1,2</sup>, Guangyu Chen,<sup>1</sup> and Liguan Guo<sup>3</sup>

<sup>1</sup>School of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou 450001, China

<sup>2</sup>Henan Province Engineering Laboratory for Information Technology of Emergency Platform, Zhengzhou 450001, China

<sup>3</sup>Henan Xinanli Security Technology Co., Ltd., Zhengzhou 450001, China

Correspondence should be addressed to Yuanpan Zheng; [ypzheng@zzuli.edu.cn](mailto:ypzheng@zzuli.edu.cn)

Received 28 March 2020; Revised 13 May 2020; Accepted 27 May 2020; Published 29 June 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Yuanpan Zheng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the change of the network communication environment in vehicular ad hoc networks (VANETs) of a smart city, vehicles may encounter security threats such as eavesdropping, positioning, and tracking, so appropriate anonymity protection is required. Based on the certificateless cryptosystem and group signature ideas, this paper proposes a certificateless group signature anonymous authentication scheme for the VANETs of a smart city. In this scheme, it can implement the process of adding, signing, verifying, and revoking group members only by simple multiplication of the elliptic curve and synchronization factor technology, which shortens the length of the signature and improves the efficiency of the signature. From the proofs of correctness and security, we know that it does not only have anonymity and traceability of the group signature scheme but also has unforgeability and forward security. According to the performance verification, this scheme has lower calculation overhead and higher authentication efficiency.

## 1. Introduction

Vehicular ad hoc networks (VANETs) [1] of a smart city, as a typical application of the Internet of Things technology, enable real-time traffic information interaction between vehicles and vehicles and between vehicles and the infrastructure. And, it has played a positive role in reducing traffic accidents and has been widely developed in the field of intelligent transportation. With the continuous change of the network environment, a variety of information security and privacy leakage issues have also emerged, seriously threatening the personal safety and personal privacy of vehicle users. Therefore, it is necessary to provide corresponding security policies, which can effectively protect the communication security and personal privacy of vehicle users while providing fast services for vehicle users.

At present, anonymous authentication technologies in VANETs mainly include PKI-based authentication, identity-based authentication, and group signature-based

authentication. In the early days, the public key infrastructure- (PKI-) based public key certificate scheme proposed by Raya and Hubaux [1] in 2007 was mainly used. This scheme requires a large number of public-private key pairings and related certificates to be stored in the vehicles. By occupying a large amount of storage space, it increases communication and computational overheads and causes certificate management problems. Shim [2] proposed an identity-based batch authentication scheme. The scheme uses a pseudonym to represent vehicle identity information and uses a pseudonym replacement strategy for each message signature to achieve message traceability. However, in this scheme, PKG knows the private keys of all users, so it is inevitable that the key escrow problem will occur.

In 1991, Chaum and Heyst [3] first proposed the concept of the group signature. It allows group members to sign anonymously on behalf of the group. The group administrator is responsible for the creation and distribution of group member keys. The group members use group member

certificates to sign on messages. The group public key is used to verify its authenticity. The verifier can only verify that the signer is from a member of the group but cannot determine the identity of specific members in the group, thereby protecting the group members' identity. In addition, the group administrator can open the signature and reveal the true identity of the signing members to resolve the dispute. But, it is computationally infeasible to distinguish whether two different group signatures come from the same signer. Therefore, the group signature technology has been widely used, and it has been gradually introduced into the anonymous authentication scheme in VANETs [4–7]. Shao et al. [5] proposed a threshold anonymous authentication protocol capable of implementing batch authentication based on the group signature. Zheng et al. [6] introduced a lightweight group signature technology, which made the group public key and signature length fixed and did not depend on the number of group members. Zhao [7] proposed a revocable group signature scheme based on the Chinese remainder theorem in VANETs. When members join and revoke, they only need to regenerate a new group public key without changing the key pairings of other members, improving the efficiency of member joining and revoking. However, in these schemes, each member needs to generate a corresponding group member certificate, which will increase storage overhead and computational overhead.

In 2003, Al-Riyami and Paterson [8] first proposed a certificateless cryptosystem. In the system, a part of the user key is provided by the key generation center and the rest is generated by the user to form the user key, which ensures that the key generation center does not know all the user's private keys, and it solves the problem of certificate management in traditional public key cryptosystems and key escrow in identity-based cryptosystems. Based on the group signature technology, Chen et al. [9] and Li et al. [10] proposed different certificateless group signature schemes. At the same time, certificateless group signature schemes applied to VANETs have also been proposed [11–17], which has also become a hotspot in the security of VANETs. Zhang et al. [12] and Chen et al. [14] used bilinear pairings to study the application of the certificateless group signature in VANETs, avoiding the problem of key escrow, without the need for certificate management, effectively reducing the system storage load.

However, the current certificateless group signature schemes are implemented with the help of bilinear pairing operations, which increases the overhead of the system operation. Therefore, this paper proposes a certificateless group signature scheme based on elliptic curves, which uses elliptic curves instead of bilinear pairings for operations. This scheme not only inherits the security and anonymity of group signature schemes but also greatly reduces the computational overhead. In particular, the introduction of the synchronization factor technology in this scheme makes it unnecessary to modify the public key information of the group administrator when the members in the group change. Only the group synchronization factor and group members' synchronization factor are calculated and

modified, which greatly reduces the calculation steps when group members join and revoke.

## 2. Preliminaries

*2.1. System Model.* In the general mode, the system model of VANETs consists of fixed RSUs (road side units) at the road side, mobile OBUs (on-board units) equipped in vehicles, and a TA (trusted authority), as shown in Figure 1.

OBUs access the VANETs through the road side deployment infrastructure RSUs and periodically broadcast their own vehicle information to other vehicles, including safety information such as the location, speed, direction, acceleration, road conditions, traffic events, and time stamps, so that other OBUs can quickly obtain useful information on the road. RSUs can broadcast and receive some signature information in the group and provide various services for the OBUs. And, when needed, they reveal the real identification of some illegal vehicles and broadcast the identification information of revoked vehicles. RSUs have their own storage space and computing capabilities. The TA, as a third-party trusted agency in this scheme, saves the real identity information of OBUs and RSUs and generates public and private key pairings of OBUs and RSUs for identification in VANETs.

*2.2. Elliptic Curve.* The elliptic curve is an encryption algorithm in the current public key encryption system, and it is also the encryption algorithm that can provide the highest encryption strength for data. The encryption strength corresponding to the encryption calculation using the 160-bit key length is equivalent to the encryption length corresponding to the RSA algorithm using the 1024-bit key length in the public key encryption system. However, the elliptic curve has the characteristics of fewer calculation parameters, shorter key length, and faster operating speed. Therefore, it is appropriate to apply the elliptic curve encryption algorithm to the VANETs with limited computing capacity, storage space, and transmission bandwidth.

*Definition 1* (elliptic curve definition). This scheme uses a 160-bit elliptical encryption algorithm. Assume that  $q$  is a large prime number and  $F_q$  is a finite field of the module  $q$ . An elliptic curve over a finite field  $F_q$  can be defined as:  $E: y^2 \equiv x^3 + ax + b \pmod{q}$ , where  $a, b, x,$  and  $y \in F_q$  and  $\Delta = 4a^3 + 27b^2 \neq 0$ .

*Definition 2* (addition of elliptic curves). Assume that the point of an elliptic curve  $P = (x_1, y_1) \in E$ ,  $-P = (x_1, -y_1)$  is the negative point of  $P$ ,  $Q = (x_2, y_2) \in E$ ,  $Q \neq -P$ , the line  $l$  passes through  $P$  and  $Q$ , and it intersects the elliptic curve at a point  $R' = (x_3, -y_3)$ , The symmetrical point about the  $x$ -axis with  $R'$  is  $R = (x_3, y_3)$  and  $R = P + Q$ . The addition cyclic group of the prime order  $q$  on the elliptic curve  $E$  is  $G_q = \{(x, y): a, b, x, y \in F_q, (x, y) \in F_q, (a, b)\}$  where  $G$  is a generator on the elliptic curve  $E$  and the scalar multiplication operation on the elliptic curve is  $kP = P + P + P + \dots + P(k, k \in Z_q^*)$ .

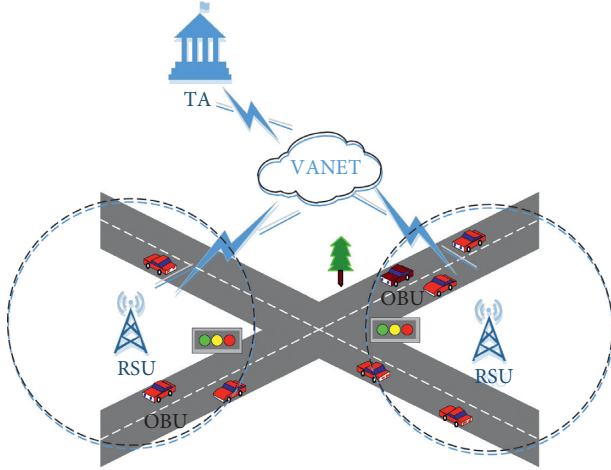


FIGURE 1: The system model of VANETs.

*Definition 3* (elliptic curve discrete logarithm problem (ECDLP)). There are two points  $P_1$  and  $P_2$  on the elliptic curve  $E$  on the finite field  $F_q$  and there exists  $k \in Z_q^*$ , such that  $P_1 = kP_2$ ; it is feasible to calculate  $P_1$  from  $k$  and  $P_2$ , but it is not advisable to calculate  $k$  from  $P_1$  and  $P_2$ .

### 3. Establishment of an Anonymous Authentication Scheme Based on Certificateless Group Signature

*Design Idea.* In this paper, the certificateless design idea is integrated into the scheme based on the group signature, which simplifies the member joining process and can resist public key replacement attacks. During the member joining process, the member  $A$  uses the private key to sign  $SK_A$ , obtains the identity signature information  $h_A$ , and sends  $(ID_A \| Y_A \| h_A \| v_A \| b_A)$  to RSU and RSU obtains  $A$ 's public key from TA to verify the identity information sent by  $A$ . It not only proves the legitimacy of  $A$  but also avoids public key replacement attacks. In addition, in the process of generating the group member certificate, the vehicle user needs to verify the identity of the group administrator RSU before accepting the member certificate to enhance the credibility of the certificate.

The certificateless group signature anonymous authentication scheme includes system initialization, public and private key generation for group administrators and group members, group member joining, signature generation, signature verification, member revocation, and opening signature. The specific work is as follows:

- (1) *System Initialization.* TA chooses the system parameters and generates the master key and its own public key, and public key information is made public.
- (2) *Public and Private Key Generation for Group Administrators and Group Members.* TA generates relevant public and private keys for administrators RSU and vehicle users OBU. The

administrator generates an initial group synchronization factor  $T$ .

- (3) *Member Joining.* The new member  $A$  joins according to the group joining method and generates a self-synchronization factor and updates the group synchronization factor.
- (4) *Signature Generation.* Group member  $A$  signs the message  $M$  based on the signature algorithm.
- (5) *Signature Verification.* In VANETs, the verifier verifies the message signature through making information and signature information public and confirms that the signed message is signed and issued by a member of the group.
- (6) *Member Revocation.* When a member in the group leaves the group for some reason, RSU recalculates the synchronization factor  $T'$  in the group according to the identity information of the member  $A$  which left the group and sends the new synchronization factor  $T'$  and related information of  $A$ 's synchronization factor to other members  $B$  in the group, which updates their synchronization factor to  $T_B$  according to the information.
- (7) *Opening Signature.* When  $A$  finds that the message signature sent by the group member vehicle user is false information or a dispute occurs between the group members, the signature is calculated by opening the signature to reveal the identity of the user.

### 4. Proposed Scheme

*4.1. Initialization.* Based on the selected security parameter  $k$ , TA generates two large prime numbers  $p$  and  $q$ , such that  $q|p-1$ . Choose the generator  $P$  on the cyclic group  $G$  on the elliptic curve of the order  $q$ . Then, choose two collision-free hash functions:  $H: \{0, 1\}^* \rightarrow Z_q^*$  and  $H_1: \{0, 1\}^* \times G \rightarrow Z_q^*$ . TA chooses a random parameter  $z \in Z_q^*$  as the system master key and calculates  $P_z = zP$  as the public key. TA makes system parameters  $\text{params} = \{p, q, G, P, P_z, H, H_1\}$  public and secretly saves the system master key  $z$ .

#### 4.2. Public and Private Key Generation

- (1) In this scheme, RSU acts as a group manager to manage vehicle members in the group. Assume that the identity information of the group manager RSU is  $ID_{RSU}$ , then RSU randomly chooses  $x_{RSU} \in Z_q^*$ , calculates  $P_{RSU} = x_{RSU}P$ , and sends  $P(ID_{RSU}, P_{RSU})$  to TA; TA randomly chooses  $r_{RSU} \in Z_q^*$ , calculates  $R_{RSU} = r_{RSU}P$  and  $S_{RSU} = r_{RSU} + zH_1(ID_{RSU} \| P_{RSU} \| R_{RSU})$ , and sends  $(R_{RSU}, s_{RSU})$  to RSU secretly, where  $R_{RSU}$  is a partial public key of RSU and  $s_{RSU}$  is a partial private key of RSU; RSU receives the information, verifies whether  $s_{RSU}P = R_{RSU} + P_zH_1(ID_{RSU} \| P_{RSU} \| R_{RSU})$  is established, and judges the validity of the partial private key  $s_{RSU}$ . At this time, RSU gets a complete private key pairing  $SK_{RSU} = (x_{RSU}, s_{RSU})$  and a complete public key

pairing  $PK_{RSU} = (x_{RSU}P, s_{RSU}P) = (P_{RSU}, S_{RSU})$ . TA saves the corresponding information  $(ID_{RSU}, P_{RSU}, S_{RSU}, s_{RSU})$  of RSU and saves the public key to the public list.

- (2) Assume that the identity information of the user  $OBU_A$  is  $ID_A$ . Through the above process, the private key pairing  $SK_A = (x_A, s_A)$  and the public key pairing  $PK_A = (P_A, S_A)$  of the user  $OBU_A$  are generated, and the public key  $PK_A$  is made public. The hash function  $H_1$  is used to generate a part of the private key.
- (3) The group manager RSU randomly chooses  $e \in Z_q^*$  and calculates  $T_0 = eP$  as the initial group synchronization factor of the group, and the engaged synchronization factor is  $T$ .

#### 4.3. Joining

- (1) When the user  $OBU_A$  wants to join the group,  $OBU_A$  randomly chooses  $y_A \in Z_q^*$  and  $b_A \in Z_q^*$  and calculates  $Y_A = y_AP$ ,  $h_A = H(ID_A \| PK_A \| Y_A \| b_A)$ , and  $v_A = y_A - h_A \cdot SK_A$ . The user  $OBU_A$  sends  $(ID_A \| Y_A \| h_A \| v_A \| b_A)$  to RSU.
- (2) RSU sends  $ID_A$  to TA, obtains  $OBU_A$ 's public key  $PK_A$ , verifies that whether  $Y_A = v_AP + h_A PK_A = Y_A$  is established, and generates a certificate for  $OBU_A$  if it holds.
- (3) RSU randomly chooses  $e_A \in Z_q^*$ , calculates  $E_A = Y_A + e_AP = (e_A + y_A)P$ ,  $h_{RSU} = H(E_A \| PK_{RSU} \| T)$ , and  $s_{RSU} = e_A + SK_{RSU} \cdot h_{RSU}$ , sends  $(E_A, h_{RSU}, s_{RSU}, T)$  to  $OBU_A$ , and stores  $(ID_A, PK_A, Y_A, b_A, E_A, E_AP, e_A, h_A, s_A)$  into the group member information list.
- (4)  $OBU_A$  verifies RSU's public key  $PK_{RSU}$  and calculates that whether  $E_A = (s_{RSU} + y_A)P - h_{RSU} PK_{RSU} = E_A$  is established. If it holds, the user  $OBU_A$  joins the group and generates the group member certificate as  $(ID_A, PK_{RSU}, Y_A, E_A, b_A, T)$ .
- (5) RSU sends  $(T, b_A)$  to other members in the group, and member  $OBU_B$  updates their synchronization factor  $T_B$ . Assuming that  $OBU_B$ 's certificate is  $(ID_B, PK_{RSU}, Y_B, E_B, b_B, T_B)$ ,  $OBU_B$  calculates a new synchronization factor as  $T_B' = T + T_B(b_B - b_B)$ , and  $OBU_B$ 's new certificate is  $(ID_B, PK_{RSU}, Y_B, E_B, b_B, T_B')$ .
- (6) RSU updates the synchronization factor as  $T' = T \cdot (b_A + x_{RSU})$ .

**4.4. Other Steps.** The remaining four steps in the scheme are, in order, signature generation, signature verification, member revocation, and signature opening.

**4.4.1. Signature Generation.** Assume that the group member  $OBU_A$  generates a signature on message  $M$ , calculates  $C_1 = E_AP + T_A PK_{RSU}$  and  $C_2 = T_AP$ ,  $C_3 = b_A E_A$ , randomly chooses  $r_1, r_2, r_3, r_4 \in Z_q^*$ , and calculates

$$d_1 = r_1 C_1 - r_2 PK_{RSU}, \quad d_2 = r_1 C_2 + r_3 S_{RSU}, \quad d_3 = r_3 P, \\ d_4 = r_3 PK_{RSU} + r_4 P, \quad c = H(PK_{RSU} \| M \| C_1 \| C_2 \| C_3 \| d_1 \| d_2 \| d_3 \| d_4), \\ s_1 = r_1 - cb_A, \quad s_2 = r_2 - cb_A T_A, \quad s_3 = r_3 - cT_A, \quad \text{and} \\ s_4 = r_4 - cE_A; \quad \text{the output signature is } RM = (c, s_1, s_2, s_3, s_4, C_1, C_2, C_3).$$

**4.4.2. Signature Verification.** The verifier calculates  $d_1' = s_1 C_1 - s_2 PK_{RSU} + c PC_3$ ,  $d_2' = s_1 C_2 + s_3 S_{RSU} + c TP$ ,  $d_3' = s_3 P + c C_2$ ,  $d_4' = c C_1 + s_4 P + s_3 PK_{RSU}$ , and  $c' = H(PK_{RSU} \| m \| C_1 \| C_2 \| C_3 \| d_1' \| d_2' \| d_3' \| d_4')$  based on  $(c, s_1, s_2, s_3, s_4, C_1, C_2, C_3)$ . If the equation  $c' = c$  holds, the verification passes.

**4.4.3. Member Revocation.** To revoke the user  $OBU_A$ , RSU calculates a new synchronization factor  $T' = T \cdot (b_A + x_{RSU})^{-1}$  based on  $(T, b_A)$ . Then, RSU sends  $(T', b_A)$  to other members in the group  $OBU_B$ , and  $OBU_B$  updates their synchronization factor  $T_B$  to  $T_B'$ , where  $T_B' = (T_B - T') \cdot (b_A - b_B)^{-1}$ .

**4.4.4. Signature Opening.** When RSU finds that the message signature sent by the group member vehicle user is false information or a dispute occurs between the group members, it calculates  $E_AP = C_1 - C_2 SK_{RSU}$  based on the signed message  $RM = (c, s_1, s_2, s_3, s_4, C_1, C_2, C_3)$  and the group manager's private key  $SK_{RSU} = (x_{RSU}, s_{RSU})$  and then finds the corresponding identity of the group member.

## 5. Anonymous Scheme Analysis

### 5.1. Correctness Analysis

**5.1.1. Correctness of Key Distribution.** After the group manager RSU receives  $(R_{RSU}, s_{RSU})$ , it verifies whether  $s_{RSU}P = R_{RSU} + P_z H_1(ID_{RSU} \| P_{RSU} \| R_{RSU})$  is established. Since  $s_{RSU}P = r_{RSU}P + zPH_1(ID_{RSU} \| P_{RSU} \| R_{RSU}) = R_{RSU} + P_z H_1(ID_{RSU} \| P_{RSU} \| R_{RSU})$ , the verification result is consistent with the result of the signature generation algorithm, so the signature scheme satisfies the correctness.

Similarly, after the user  $OBU_A$  receives  $OBU_A$ , it verifies whether  $s_AP = R_A + P_z H_1(ID_A \| P_A \| R_A)$  is established. Since  $s_AP = r_AP + zPH_1(ID_A \| P_A \| R_A) = R_A + P_z H_1(ID_A \| P_A \| R_A)$ , the signature scheme satisfies the correctness.

**5.1.2. Correctness of Signature in Joining.** After RSU receives the signature information  $(ID_A \| Y_A \| h_A \| v_A \| b_A)$  from the user  $OBU_A$ , if  $(h_A, v_A)$  is a legitimate signature, the equation  $Y_A = v_AP + h_A PK_A = y_AP - h_A SK_A P + h_A PK_A = Y_A$  holds, and then RSU calculates  $h_A' = H(ID_A \| PK_A \| Y_A \| b_A)$  based on  $Y_A$  and gets  $h_A' = h_A$ . And so, the  $r_4 P - cE_AP + r_3 PK_{RSU}$  signature is valid, that is, the identity of the user  $OBU_A$  is valid.

Similarly, when  $OBU_A$  receives the message  $(E_A, h_{RSU}, s_{RSU}, T)$  sent by RSU and calculates  $E_A = (s_{RSU} + y_A)P - h_{RSU} PK_{RSU}$  based on RSU's public key  $PK_{RSU}$  and

$s_{RSU} = e_A + SK_{RSU} \cdot h_{RSU}$ , then the equation  $E'_A = E_A$  holds. And so, the signature is valid.

**5.1.3. Correctness of Group Signature.** If  $(c, s_1, s_2, s_3, s_4, C_1, C_2, C_3)$  is a legitimate signature, the verifier calculates  $d'_1 = s_1 C_1 - s_2 PK_{RSU} + c PC_3 = r_1 C_1 - cb_A E_A P - cb_A T_A PK_{RSU} - r_2 PK_{RSU} + cb_A T_A PK_{RSU} + cb_A E_A P = r_1 C_1 - r_2 PK_{RSU} = d_1$ ,  $d'_2 = s_1 C_2 + s_3 S_{RSU} + c TP = r_1 C_2 - cb_A T_A P + r_3 S_{RSU} - c T_A S_{RSU} + c TP = r_1 C_2 + r_3 S_{RSU} = d_2$ ,  $d'_3 = s_3 P + c C_2 = r_3 P - c T_A P + c T_A P = r_3 P = d_3$ , and  $d'_4 = c C_1 + s_4 P + s_3 PK_{RSU} = c E_A P + c T_A PK_{RSU} + r_4 P - c E_A P + r_3 PK_{RSU} - c T_A PK_{RSU} = r_3 PK_{RSU} + r_4 P = d_4$ , based on  $TP = T_A (b_A + x_{RSU})P = T_A b_A P + T_A S_{RSU}$  and gets  $c t = c$  from the existing public information, so the signature verification algorithm is correct.

**5.2. Unforgeability.** Unforgeability means that the group certificate of the members in the group is unforgeable.

In this scheme, RSU's private key pairing is  $SK_{RSU} = (x_{RSU}, s_{RSU})$ , where  $s_{RSU} = r_{RSU} + z H_1(ID_{RSU} \| P_{RSU} \| R_{RSU})$ ; the group certificate for the group member  $OBU_A$  is  $(ID_A, PK_{RSU}, Y_A, E_A, b_A, T_A)$ , where  $E_A = Y_A + e_A P = (e_A + y_A)P$ ,  $Y_A = y_A P$ , and the synchronization factor of the group  $T$  and the synchronization factor of the group member  $OBU_A$  have the following relationship:  $T = T_A (b_A + x_{RSU})$ .  $y_A, b_A, x_{RSU}$ , and  $e_A$  are private to group members  $OBU_A$  and RSU, respectively, so no single party can complete the group member certificate creation independently. Therefore, the group certificate is unforgeable.

**5.3. Forward Security.** When group member  $OBU_A$  joins the group, the group synchronization factor  $T$  is updated as follows:  $T' = T \cdot (b_A + x_{RSU})$ , based on  $b_A$  provided by  $OBU_A$ , and the synchronization factors of other members  $OBU_B$  in the group are updated as follows:  $T'_B = T + T_B (b_B - b_B)$ ; when the group member  $OBU_A$  is revoked, the group synchronization factor  $T$  is updated as follows:  $T' = T \cdot (b_A + x_{RSU})^{-1}$ , and the synchronization factors of other members  $OBU_B$  in the group are updated as follows:  $T'_B = (T_B - T) \cdot (b_A - b_B)^{-1}$ . It can be seen that the signature in the verification phase and the synchronization factor used in the verification phase will be updated synchronously according to the membership addition and revocation. After the update, the previous signature verification equation will not be established, so the forward security can be guaranteed.

**5.4. Performance Analysis.** In this section, performance analysis will be performed in terms of communication costs and calculation costs. For this scheme, the communication cost needs to consider the length of the group manager's public key and the length of the group member's signature. In the calculation aspect, the cost of joining the group, the cost of revoking the group, the cost of computing the signature, and the cost of verifying the signature are considered. Compared with other group signature schemes, some

performance analysis comparisons are made as given in Table 1, where  $N$  represents the number of current group members and the number of joined and revoked members each time is set to 1.

In this scheme, the length of the group manager's public key and the length of the group member's signature information are not directly related to the number of members in the group and are constant.

In this scheme, when joining and revoking, the synchronization factor of each user needs to be updated, so the cost of joining and revoking is  $O(N)$ .

In this scheme, the efficiency of the calculation cost of the information signature and the verification cost of the signature information are both constant, and the number of group members does not affect the time spent on signature and verification.

For this scheme, the performance analysis mainly considers the cost of group membership joining and revocation, the cost of information signature, and the cost of verifying signature information.

According to the literature [15], we choose a hardware platform consisting of Intel I7-6700 and Windows7 with 8G processor memory. By performing elliptic curve/bilinear pairing simulation experiments multiple times and taking the average value of the results, the operation execution schedule can be obtained as shown in Table 2. The comparison of this paper's average execution time of simulation operations is shown in Figure 2.

Considering the overall performance of the scheme, we will focus on analyzing the time overhead in the signature generation and signature verification process. This scheme is compared with the existing schemes [14, 15]. In the signature generation phase, scalar multiplication of bilinear pairs is mainly used in the scheme [14, 15]. The overall multiplication operation is less than this scheme, but the length of a single multiplication operation is longer than the elliptic curve multiplication and modular multiplication operations used in this scheme, and the overall time overhead is greater than the time overhead of this scheme; moreover, in the signature generation, the calculation of  $2T_{EC\_MUL} + 2T_{MUL}$  is a fixed calculation, and it does not need to participate in each calculation process, which can further reduce the calculation cost of group members when performing signature generation. In the signature verification phase, the time-consuming bilinear operation in the scheme [14, 15] increases the time overhead, and the signature verification process of this scheme is not much different from the signature generation calculation overheads, as shown in Table 3. The comparison of signature generation and signature verification overhead for the three schemes is shown in Figure 3.

In the process of the group member joining, since the group members and the group management need to verify the identity of each other, the group members need to perform four elliptic curve multiplication operations and two hash comparisons. During the joining and revocation stages of group members, the group management broadcasts the synchronization coefficients of new members, and the members within the group update their respective

TABLE 1: Performance analysis.

| Scheme      | Length of public key | Signature length | Joining cost | Revocation cost | Signature cost | Verification cost |
|-------------|----------------------|------------------|--------------|-----------------|----------------|-------------------|
| LPY [4]     | $O(1)$               | $O(1)$           | $O(N)$       | $O(1)$          | $O(1)$         | $O(1)$            |
| YJD [11]    | $O(1)$               | $O(1)$           | $O(N)$       | $O(1)$          | $O(N \log n)$  | $O(1)$            |
| This scheme | $O(1)$               | $O(1)$           | $O(N)$       | $O(N)$          | $O(1)$         | $O(1)$            |

TABLE 2: Average execution time of simulation operations.

| Symbol        | Description                       | Execution time (ms) |
|---------------|-----------------------------------|---------------------|
| $T_{MC\_MUL}$ | Multiplication on elliptic curves | 0.3476              |
| $T_{MC\_ADD}$ | Addition on elliptic curves       | 0.002               |
| $T_{MUL}$     | Modular multiplication            | 0.0119              |
| $T_H$         | General hash function operations  | 0.0012              |
| $T_{PB\_SM}$  | Scalar multiplication             | 0.817               |
| $T_{PB}$      | Bilinear operation                | 5.5852              |

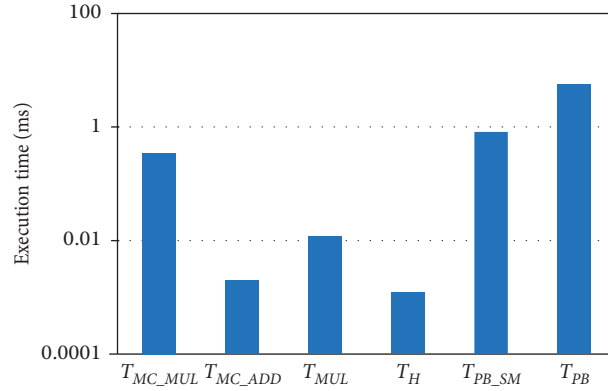


FIGURE 2: Comparison of average execution time of simulation operations.

TABLE 3: Signature generation and verification calculation overhead.

| Scheme      | Signature generation             |                    | Signature verification          |                    |
|-------------|----------------------------------|--------------------|---------------------------------|--------------------|
|             | Calculation overhead             | Time overhead (ms) | Calculation overhead            | Time overhead (ms) |
| Scheme [14] | $2T_{PB} + 1T_H + 10T_{PB\_SM}$  | 19.3416            | $2T_{PB} + 1T_{PB\_SM}$         | 11.9874            |
| Scheme [15] | $5T_{PB\_SM}$                    | 4.085              | $4T_{PB} + 2T_{PB\_SM}$         | 23.9748            |
| This scheme | $4T_{MC\_MUL} + 12T_{MUL} + T_H$ | 1.5344             | $4T_{MC\_MUL} + 7T_{MUL} + T_H$ | 1.4749             |

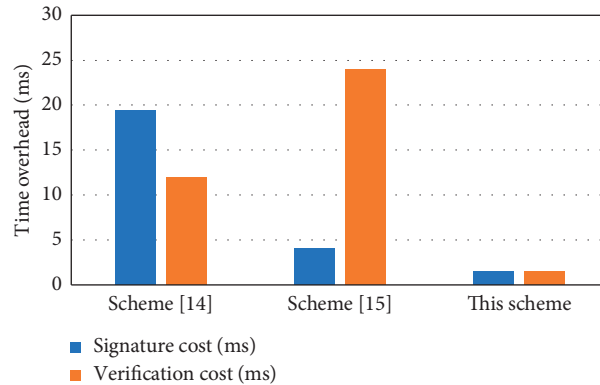


FIGURE 3: Comparison of signature generation and signature verification overhead.

synchronization factors. Without modifying the group public key, the calculation costs caused by changes in the members of the group will be spent, allocating sales to members in the group and reducing the calculation requirements for group management.

## 6. Conclusion

Aiming at the problem of low authentication efficiency in the anonymous authentication scheme in VANETs, this paper proposes a certificateless elliptic curve anonymous authentication scheme. Though based on a certificateless signature scheme, this scheme does not have to consider certificate maintenance and key escrow issues. It also uses elliptic curves to perform calculations on the basis of certificatelessness and introduces synchronization factor technology to further improve computing efficiency of group members when joining, revoking, and signing. The analysis of the scheme shows that the proposed scheme can not only ensure the anonymity and traceability of the group signature scheme but also ensure unforgeability and forward security under the premise of correctness. The partial key generation scheme adopted in this scheme effectively ensures the security of user keys, and there is no need to save too much certificate information in the system, and the calculation and storage overhead is low. Therefore, it is very suitable for OBUs and RSUs with very limited computing and storage space in the VANETs.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the Natural Science Foundation of China (grant no. 51404216) and the Henan Province Programs for Science and Technology Development (grant nos. 202102210180, 172102310670, and 152102310374).

## References

- [1] M. Raya and J.-P. Hubaux, "Securing vehicular ad hoc networks," *Journal of Computer Security*, vol. 15, no. 1, pp. 39–68, 2007.
- [2] K.-A. Shim, "Reconstruction of a secure authentication scheme for vehicular ad hoc networks using a binary authentication tree," *IEEE Transactions on Wireless Communications*, vol. 12, no. 11, pp. 5386–5393, 2013.
- [3] D. Chaum and V. E. Heyst, "Group signatures," *Advances in Cryptology—EUROCRYPT'91*, Springer, Berlin, Germany, pp. 257–265, 1991.
- [4] C. I. Fan, W. Z. Sun, S. W. Huang, W. Juang, and J. Huang, "Strongly privacy-preserving communication protocol for VANETs," in *Proceedings of the 2014 Ninth Asia Joint Conference on Information Security*, IEEE, Wuhan, China, September 2014.
- [5] J. Shao, X. Lin, R. Lu, and C. Zuo, "A threshold anonymous authentication protocol for VANETs," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 3, pp. 1711–1720, 2016.
- [6] M. Zheng, Y. Duan, and H. Lyu, "Research on identity authentication protocol group signature-based in Internet of vehicles," *Advanced Engineering Sciences*, vol. 50, no. 4, pp. 130–134, 2018.
- [7] Z. Zhao, *Reserrch on Efficient Group Signatures Schemes in VANET*, Xidian University, Xi'an, China, 2015.
- [8] S. S. Al-Riyami and K. G. Paterson, "Certificateless public key cryptography," *Advances in Cryptology—ASIACRYPT 2003*, Springer, Berlin, Germany, pp. 452–473, 2003.
- [9] H. Chen, C. Zhu, and R. Song, "Journal of computer research and development," *Journal of Computer Research and Development*, vol. 47, no. 2, pp. 231–237, 2010.
- [10] F. Li, P. Liu, and Z. Zhu, "Certificateless signature and group signature schemes based on bilinear pairings," *Computer Engineering*, vol. 37, no. 24, pp. 18–21, 2011.
- [11] J. Yin, *The Research on Certificateless Authenticated Group Key Management in Ad Hoc Network*, Beijing Institute of Technology, Beijing, China, 2016.
- [12] X. Zhang, Y. Xu, and J. Cui, "Anonymous authentication protocol based on certificateless signature for vehicular network," *Computer Engineering*, vol. 42, no. 3, pp. 18–28, 2016.
- [13] C. Song, M. Zhang, W. Peng, Z. Jia, Z. Liu, and X. Yan, "Research on pairing-free certificateless batch anonymous authentication scheme for VANET," *Journal on Communications*, vol. 38, no. 11, pp. 35–43, 2017.
- [14] Y. Chen, X. Cheng, S. Wang, and M. Gao, "Research on certificateless group signature scheme based on bilinear pairings," *Netinfo Security*, vol. 3, pp. 53–58, 2017.
- [15] N. Zhao, G. Zhang, and X. Gu, "Certificateless aggregate signature scheme for privacy protection in VANET," *Computer Engineering*, vol. 46, no. 1, pp. 114–128, 2020.
- [16] Y. Gan, K. Wang, and L. He, "RFID tag dynamic ownership transfer protocol of multi-owner with TTP weight," *Journal of Light Industry*, vol. 33, no. 1, pp. 72–78, 2018.
- [17] Y. Xiao, J. Du, M. Wen, K. Zhou, J. Jiao, and J. Pei, "Traffic sign detection and recognition based on color features and improved support vector machine algorithm," *Journal of Light Industry*, vol. 33, no. 3, pp. 57–65, 2018.

## Research Article

# Adaptive Language Processing Based on Deep Learning in Cloud Computing Platform

Wenbin Xu<sup>1</sup> and Chengbo Yin<sup>2</sup>

<sup>1</sup>Department of English Language and Literature, China University of Petroleum (East China), Qingdao, Shandong 266580, China

<sup>2</sup>School of Data Science, Qingdao Huanghai University, Qingdao 266427, Shandong, China

Correspondence should be addressed to Wenbin Xu; [xuwenbin@upc.edu.cn](mailto:xuwenbin@upc.edu.cn)

Received 28 March 2020; Revised 7 May 2020; Accepted 27 May 2020; Published 19 June 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Wenbin Xu and Chengbo Yin. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the continuous advancement of technology, the amount of information and knowledge disseminated on the Internet every day has been developing several times. At the same time, a large amount of bilingual data has also been produced in the real world. These data are undoubtedly a great asset for statistical machine translation research. Based on the dual-sentence quality corpus screening, two corpus screening strategies are proposed first, based on the double-sentence pair length ratio method and the word-based alignment information method. The innovation of these two methods is that no additional linguistic resources such as bilingual dictionary and syntactic analyzer are needed as auxiliary. No manual intervention is required, and the poor quality sentence pairs can be automatically selected and can be applied to any language pair. Secondly, a domain adaptive method based on massive corpus is proposed. The method based on massive corpus utilizes massive corpus mechanism to carry out multidomain automatic model migration. In this domain, each domain learns the intradomain model independently, and different domains share the same general model. Through the method of massive corpus, these models can be combined and adjusted to make the model learning more accurate. Finally, the adaptive method of massive corpus filtering and statistical machine translation based on cloud platform is verified. Experiments show that both methods have good effects and can effectively improve the translation quality of statistical machines.

## 1. Introduction

Currently, corpus-based translation system relies on large-scale bilingual parallel corpus, uses the translation model to estimate the probability, and selects the final translation result based on the translation probability. The advantage of the corpus-based translation method over the rule-based translation method is that it does not require much human and material participation in the construction of the model. The researchers themselves do not need to master the level of linguistic experts in the mastery of the two languages. The threshold is not so high, which allows more interested scholars and researchers to invest in it. Depending on the specific translation strategy, corpus-based machine translation can be divided into statistical-based machine

translation and instance-based machine translation. The statistical-based method is the mainstream method of current machine translation.

The early stages of statistical machine translation development use only some coarse-grained features, such as bidirectional phrase translation probabilities [1–3], bidirectional lexical translation probabilities [4], vocabulary length penalties [5], phrase lengths, punishment [6], language model [7], and sequence model [8–10]. Many systems use only these 10–20 features to complete the translation process and use the minimum error rate training (MERT) method [11–13] to perform feature weight adjustment. With the development of statistical translation models and the widespread use of massive data, researchers have found that the use of fine-grained



features [14] can further improve the accuracy of the translation system. However, the use of a large number of fine-grained features poses a great challenge to the adjustment of feature weights. The traditional MERT method can only adjust the weights of dozens of features but cannot do anything for a translation system with thousands of features. References [15–17] proposed a training algorithm based on max-violation perceptron and forced decoding [18], which can be used to translate the system by using all bilingual training data, large-scale discriminative training, and support for tens of millions of sparse features. Compared to the MERT and PRO methods, this approach can bring very significant performance improvements [19, 20] and further maximizes the use of perceptual machine training methods. The hierarchical phrase translation system has also achieved good results. The traditional statistical machine translation domain adaptive method usually migrates the model for a single domain. For example, the training data is news corpus, and the test data is network corpus. However, most practically in the application scenario, it is necessary to perform model migration on multiple domains at the same time. For example, for online translation services, the user’s input is usually text from various fields, which requires the statistical machine translation model to process automatically according to the actual input. The field adaptive research of deep learning translation is still relatively few, and the existing work has not given a clear domain label. However, the actual translation of scientific and technological literature often faces multiple professional fields, and the use of existing knowledge to organize information, such as the keywords of the paper, the scientific and technological word system, and other knowledge to obtain more clear semantic tags, helps to divide the corpus more finely.

In view of this, this paper mainly studies the multidomain adaptive method of statistical machine translation based on massive corpus under the cloud computing platform. Firstly, two corpus screening strategies are proposed, based on the double-sentence pair length ratio method and the word alignment information based method. The innovation of these two methods is that no additional linguistic resources such as bilingual dictionary and syntactic analyzer are needed as auxiliary. No manual intervention is required, and the poor quality sentence pairs can be automatically selected, and can be applied to any language pair. Secondly, a domain adaptive method based on massive corpus is proposed. The method based on massive corpus utilizes massive corpus mechanism to carry out multidomain automatic model migration. In this domain, each domain learns the intradomain model independently, and different domains share the same general model. Through the method of massive corpus, these models can be combined and adjusted to make the model learning more accurate. Finally, the adaptive method of massive corpus filtering and statistical machine translation based on cloud platform is verified. Experiments show that both methods have good effects and can effectively improve the translation quality of statistical machines.

## 2. Massive Corpus Screening Strategy under Cloud Computing Platform

*2.1. Cloud Computing Platform Framework.* The Hadoop Distributed File System (HDFS) can be deployed on a large number of inexpensive machines to store up terabytes and petabytes of data in a highly fault-tolerant and reliable manner. It combines well with the MapReduce model to provide high-throughput data access. The structure of DFS is shown in Figure 1.

As can be seen in Figure 1, an HDFS cluster that consists with a NameNode and multiple DataNodes was discussed. The metadata and the DataNode are actual data. The application accesses the NameNode to get the metadata of the file, and the actual I/O operation is directly interacting with the DataNode. The NameNode is the primary control server responsible for managing the file system namespace and coordinating application access to files, recording any changes to the namespace or changes to their properties. The DataNode is responsible for storage management on the physical node where the file is located. The feature of HDFS is that the data is “write once, read many times.” The files of HDFS are generally divided into different data blocks according to a certain size, and each data block is dispersed into different DataNodes as much as possible. In addition to completing the namespace operation of the file system, the NameNode also determines the mapping of the data block to the DataNode.

*2.2. Massive Corpus Screening Strategy.* For statistical machine translation systems, the intuitive understanding is that increasing the size of the training data can help improve system performance. Massive data is easier to obtain in today’s information environment than ever before. Scholars have built knowledge bases such as parallel sentence pairs and bilingual dictionaries by crawling bilingual web pages [21]. There are more and more sources of corpora, from multilingual websites, comparable bilingual corpora, human translated text, and more. The scale of building parallel corpora has been large, and it can be used for statistical machine translation system training. Too many errors must affect statistical machine translation systems that rely on data quality. In view of the fact that there is no qualitative change in the current statistical model, it is necessary to acquire the model features by training the corpus. Therefore, in order to train a high-performance statistical machine translation system, it is necessary to process and screen the training data. In this paper, two methods are used to filter the noise sentence pairs in the bilingual parallel corpus: the method based on the sentence pair length ratio and the method based on the alignment information.

*2.2.1. Method Based on Sentence to Length Ratio.* In general, the length of a pair of statements that are translated should be proportional to a certain ratio. However, most parallel corpora contain sentence pairs that do not match the length ratio. These sentence pairs are usually noise in the corpus. Noise phenomena caused by the length ratio include

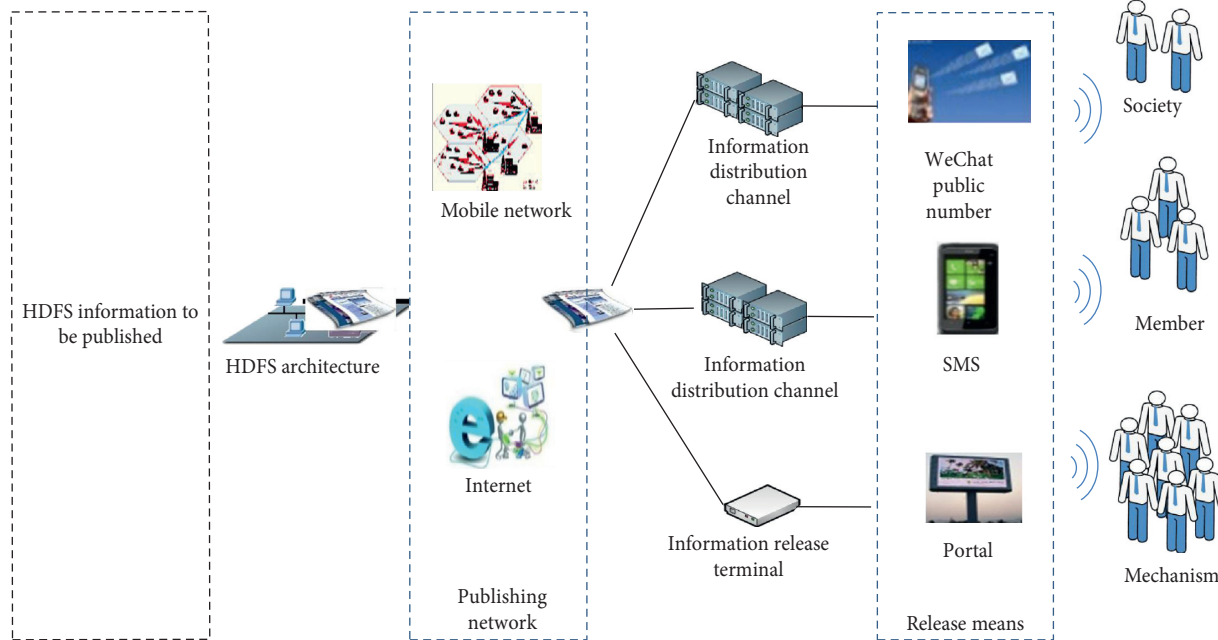


FIGURE 1: HDFS and DFS structure diagram.

monolingual errors, alignment errors, and inclusion of unknown tags (html tags, etc.). These phenomena have been observed, and many noise sentence pairs whose length ratios do not conform to the regularity are found in the experimental corpus. Some examples are listed in Table 1.

In order to remove such pairs of erroneous sentences, we set a length rule that defines the length ratio as

$$LR(f, e) = \frac{|f|}{|e|}, \tag{1}$$

where  $f$  is the source sentence,  $e$  is the target sentence, and  $|f|$  and  $|e|$  are the number of words after the source and target sentences.

The method based on the length ratio is usually based on linguistic knowledge, and the artificially set length is lower than the upper and lower limits. This paper assumes that the noise sentence in the corpus is less than the normal sentence pair; that is to say, there is a continuous range of ratios, which is the majority of the normal sentence pairs in this range. Therefore, the threshold is set according to the statistical distribution characteristics of the length ratio; that is, the sentence pairs whose length is less than the total number are filtered out. This has the advantage that thresholds can be set for different language pairs without the need for specific linguistic knowledge.

**2.2.2. Method Based on Word Alignment.** The word alignment problem is the task of finding the alignment of words in a given two-state pair. It is a key step in statistical machine translation. The word alignment model has been studied for a long time, and people use different methods for bilingual word alignment. Run the IBM model from both directions and merge the results of the two word alignments. In general, the intersection contains relatively reliable

TABLE 1: Noise sentence pair example.

| Sentence example  | Error                                 |
|---|---------------------------------------|
| (of an enterprise, institution, etc.)<br>Shift to another system                        | Sentence level alignment is incorrect |
| Falco jigger &bra; lagger falcon &ket;<br>c & w &bar; c-and-w &bar; country-and-western | Source unknown tag<br>Unknown label   |
| 1. (of a liquid) to settle, 2. clear;<br>transparent                                    | Sentence level alignment is incorrect |
| Take cognizance of  | Incorrect content                     |

alignment points; that is, the alignment point is highly accurate but does not contain all reliable alignment points; and the assembly contains most of the desired alignment points; that is, the recall rate is high but introduces additional errors. A good alignment point is adjacent to other good alignment points. Therefore, the algorithm starts from aligning the intersections. In the expansion step, adjacent alignment points located in the union but not in the intersection are added, and finally points that are not aligned in both directions are added. The pseudocode for this algorithm is given in Table 2.

The position of the two sentences in the corpus is adjacent. The occurrence of this situation is the automatic noise extraction of the parallel sentence to the technology, because it is impossible to judge the correct alignment sentence pair (the correct alignment sentence corresponds to <discountant opinions, discordant opinions>, or <discondant opinions> which is the second sentence in the table). A similar situation has occurred many times in the corpus we use.

Figure 2 shows the alignment matrix of the two sets of sentence pairs. As shown in Figure 2, it shows the results of two unidirectional alignments in English and Chinese, and

TABLE 2: Pseudocode for grow-diag-and-final heuristic word alignment extension.

---

```

GROW DIAG-AND-FINAL( $e2f$ ):
  neighboring = ((-1, 0), (0, -1), (1, 0), (0, 1),
                (-1, -1), (-1, 1), (1, -1), (1, 1))
  alignment = intersect( $e2f$ ,  $f2e$ );
  GROW-DIAGQ; FINAL( $e2$ ); FINAL( $f2$ );
  GROW DIAL-ANDQ:
    iterate until no new points added
      for English word  $e=0, 1, 2 \dots en$ 
        for Foreign word  $f=0, 1, 2 \dots fh$ 
          if (e aligned with f)
            for each neighboring point( $e$ -new,  $f$ -new):
              if( $e$ -new not aligned or  $f$ -new not aligned)and
                ( $e$ -new,  $f$ -new)in union( $e2f$ ,  $f2e$ )
                add alignment point( $e$ -new,  $f$ -new)
  FINAL(a):
    for English word  $e$ -new = 0, 1, 2 ... en
      for foreign word  $f$ -new = 0 ... fn
        if( $e$ -new not aligned or  $f$ -new not aligned)and
          ( $e$ -new,  $f$ -new)in union( $e2f$ ,  $f2e$ )
        add alignment point( $e$ -new,  $f$ -new)

```

---

the bidirectional alignment matrix on the right is obtained by the grow-diag-and-final algorithm. It can be seen that the intersection of two unidirectional alignments is only a matter of discretion, discordant; that is to say, when the grow-diag-and-final extension is performed, there is only one alignment result that is originally considered reliable. After the expansion, the result of the obvious error alignment is obtained. This error not only affects the alignment quality of itself, but also affects the rule extraction result of the translation system. For example, in the phrase system, the above sentence pairs are extracted from the rules of discriminating, inconsistent. This kind of rule does not play any role in translation decoding; even if the decoder selects the rule, it will only reduce the translation quality. Therefore, similar problems should be avoided as much as to improve the quality of the translation or to reduce the size of the rule set.

To this end, we propose a sentence-pair filtering method based on the grow-diag-and-final extension method. We consider expanding the number of alignment results EC and the number of alignment results of the intersection alignment IC. When the extended alignment result exceeds the intersection alignment result by a certain amount, we think that the alignment result is unreliable. We set the filtering rules based on the word alignment extension and use the following to judge whether the word alignment results are reliable:

$$ER(f, e) = \frac{IC - EC}{UC}. \quad (2)$$

### 3. Statistical Machine Translation Adaptation Based on Massive Corpus

*3.1. Statistical Machine Translation Adaptation.* This section introduces a domain adaptive approach based on massive corpus. The main idea is that the training of our model is

mainly divided into three steps: first, selecting the data in the domain is according to the predefined domain; second, training the domain model and the general model is to construct the statistical machine translation system; third, using massive corpus technology makes joint adjustments to multiple domain systems.

According to the above, the first step in this work is to select the in-domain bilingual control data from all the bilingual training data to train the translation model. Since the monolingual data in a specific field can be obtained in large quantities, we draw on the method of bilingual cross section data selection [22] to obtain bilingual data in the field:

$$[H_{I\text{-src}}(s) - H_{G\text{-src}}(s)] + [H_{I\text{-tgt}}(t) - H_{G\text{-tgt}}(t)]. \quad (3)$$

This bilingual cross-entropy-based criterion tends to choose a sentence pair that is more similar to the data distribution in the domain but different from the general data distribution. Therefore, this method considers that the sentence pair with larger cross-entropy difference should be selected.

In the second step, we use the training data in the selected domain to build a statistical machine translation system based on the hybrid model. Specifically, we adopted the idea of a hybrid model to build  $N$  machine translation systems for  $N$  predefined fields; each of which is a log-linear model. For each system, the optimal translation result  $f$  is given by

$$f = \arg \max_r \{P(f | e)\}. \quad (4)$$

For each machine translation system, two translation models and two language models are included. The translation model of a specific field is trained by the bilingual data selected by the data selection method introduced in the previous section, and the translation model of the general domain is trained using all bilingual data. For the language model, we reuse the language-specific and general-language models of the specific domain trained for data selection in the previous section. Compared to a translation system that does not do domain migration, this system with a hybrid model can better balance the general translation knowledge and domain-specific translation knowledge and can benefit from two aspects.

In the third step, it is necessary to adjust the feature weights in different machine translation systems. The traditional method of arranging is generally directed to a single system. The method described in this section regards translation systems in different fields as related translation tasks, and joints are coordinated under the framework of massive corpus. There are two reasons for using massive corpora:

- (1) The translation system of a specific domain shares the same general domain translation model and language model, and the massive corpus mechanism can make better use of the common translation knowledge of translation tasks in different fields.
- (2) By forcing the general domain translation model and the language model to behave the same in different

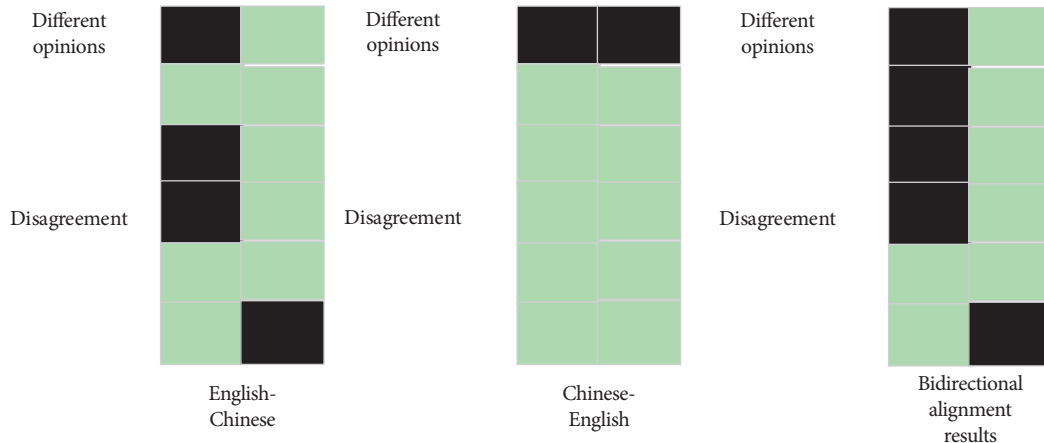


FIGURE 2: Schematic diagram of the alignment matrix of the wrong sentence pair.

fields, massive corpus provides a regularization mechanism to prevent model overfitting. Formally, the objective function of using massive corpus to adjust parameters is represented by the following formula:

$$\min_w \left\{ \sum_{i=1}^N \text{Loss}(E_i, e(F_i, w_i)) \right\}. \quad (5)$$

In order to be able to efficiently coordinate the parameters, we have improved an asynchronous stochastic gradient descent algorithm to optimize and borrowed the idea of pairwise ranking to use the perceptron algorithm to update the feature weights.

$$L(w_i) = -(w_i, v^1 - v^2). \quad (6)$$

We first use the machine translation system to generate the  $N$  best translation result candidates ( $N$ -best), which are reordered and combined into pairs by scoring with smooth sentence level BLEU. Specifically, similar to the asynchronous gradient descent algorithm, we divide the  $N$  best translation result candidates into three parts: the best 10% (high), the middle 80% (middle), and the worst 10% (low). These three parts of the translation result candidates are used for two-two sorting, in which we choose “high one,” “medium one low,” and “high one low” to combine in pairs, but will not select two of the same part Candidate combinations that are paired. The basic idea of constructing a sample in this way is that the algorithm can better have the discriminability of distinguishing between high quality and low quality translation results.

**3.2. Neural Network Deep Fusion Model.** The algorithm based on domain knowledge uses the explicit discrete features of domain knowledge, and the deep labeling algorithm uses the hidden continuous features of deep learning. The sentence domain probability vectors obtained by the two methods are different.

Combining the domain labeling algorithm based on domain knowledge and the domain labeler based on deep learning, a multilayer perceptron based on the top layer is designed as a deep fusion model of the neural network. The architecture is shown in Figure 3. The preprocessing of the sentence to be labeled is mainly word segmentation and garbled filtering. The preprocessed results are input to the knowledge-based domain tagger and deep learning-based domain tagger to obtain the domain knowledge-based probability vector and probability vectors for deep learning. The top-level neural network deep fusion model is a two-layer perceptron, and the hidden layer is two receiving four-dimensional vectors. Neuron, the activation function, is set to the ReLU (Rectified Linear Unit) function.

$$\begin{aligned} y_1 &= \text{ReQ}(W_1 x_1 + b_1), \\ y_2 &= \text{ReQ}(W_2 x_2 + b_2). \end{aligned} \quad (7)$$

The deep mixed neural model obtained through this fusion will well combine explicit and invisible knowledge, merging the advantages of discrete features and continuous features, and make the probability vector and decision category of each sentence more accurate. Thereby, the adaptation problem in the field of machine translation is better improved, and, for the data to be translated in a specific field, a higher quality translation output will be obtained.

## 4. Experiments and Results

**4.1. Massive Corpus Screening Experiment Verification.** The experimental part of this paper runs on a separate server. The specific software and hardware configuration is shown in Table 3. Because Hadoop installation is a stand-alone mode, the comparison and analysis of experimental results focus on the impact of the proposed method on translation quality.

Under the cloud translation platform, bilingual parallel corpora are a wide range of sources, such as translated manuscripts completed by translators, officially published bilingual materials, and automatic extraction of multilingual web pages. The quality of corpus is uneven. Therefore, in order to test whether the method is effective, the training

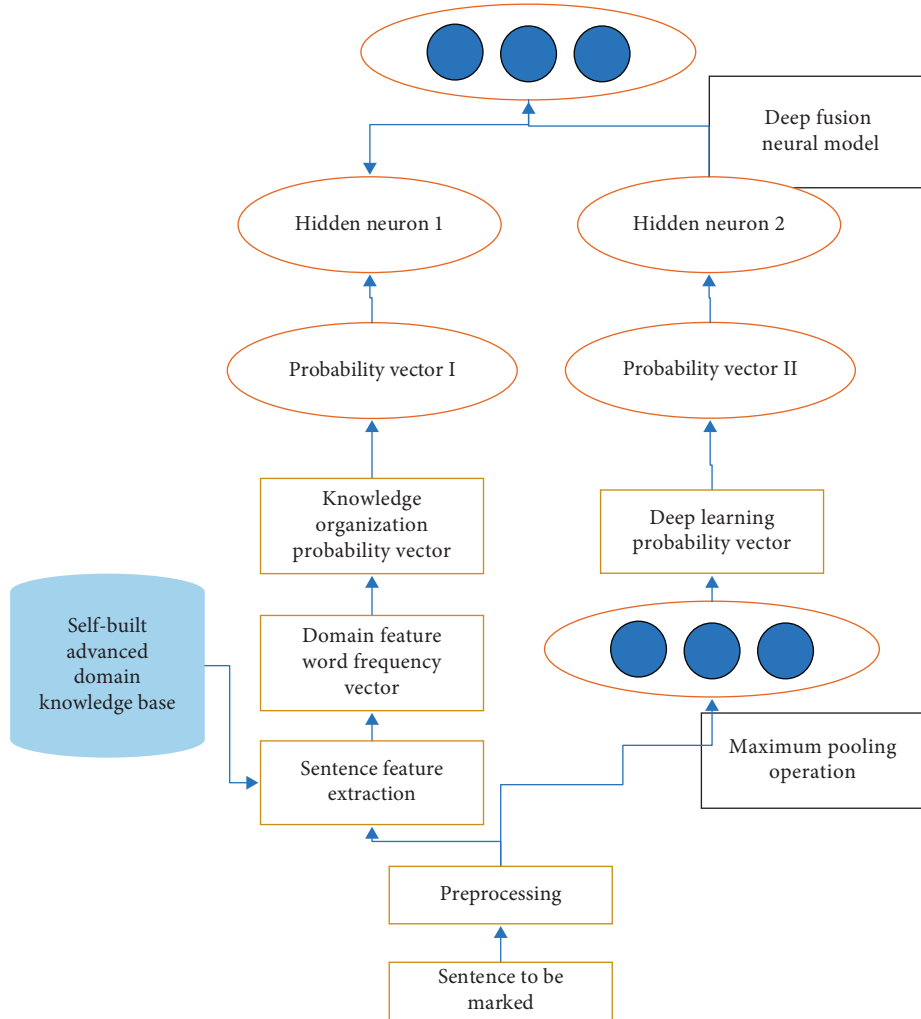


FIGURE 3: Neural network deep fusion model framework.

TABLE 3: Experimental environment.

| CPU                           | RAM (GB) | Operating system |
|-------------------------------|----------|------------------|
| Intel (R)-Xeon (R) 11 2.93 Hz | 96       | Ubuntu 12.04.1   |

corpus selected is mostly from the network. It contains 1,937,289 bilingual parallel sentences as a training set, which mixes multiple fields of content. The English language model uses the Xinhua part of LDC2007T07, the test set, and the development set. The test set contains a British news review in cwmt2011, and nist02 to nist08. The data sets are specifically listed in Table 4.

The tools used in the experiment include open source word alignment tool GIZA ten +, open source statistical machine translation system Moses, open source language model training tool irstlm, and Chinese word segmentation tool icclas. The translation model used is the phrase model, and the parameters use the standard settings of Moses.

First, we count the distribution of the length ratio of the English-Chinese sentence in the training set. The result is shown in Figure 4. The ordinate indicates the number of sentence pairs, and the abscissa indicates the ratio of the

TABLE 4: Source of experimental data.

| Corpus type  | Corpus name             | Corpus size             |
|--------------|-------------------------|-------------------------|
| Training set | web.ch-en               | 1,937,289 sentence pair |
|              |                         |                         |
| Test set     | cwmt2011                | 1006 sentence pair      |
|              | nist02                  | 878 sentence pair       |
|              | nist03                  | 919 sentence pair       |
|              | nist04                  | 1788 sentence pair      |
|              | nist05                  | 1082 sentence pair      |
|              | nist06                  | 1664 sentence pair      |
|              | nist08                  | 1357 sentence pair      |
| LDC2007T07   | 9,685,593 sentence pair |                         |

length of the sentence between English and Chinese. We can find the ratio of the length of the sentence to a certain distribution law. When the ratio is 1.0, the sentence number is the most, and there are 297,341 pairs of sentences. The figure shows that the highest point of the ratio (1.0) is also relatively large in number of sentences. This verifies our hypothesis that the length ratios of the two languages conform to the law in a continuous range.

As shown in the above figure, the contrast value of the sentences appearing in the corpus is 0.1–66.0 due to the

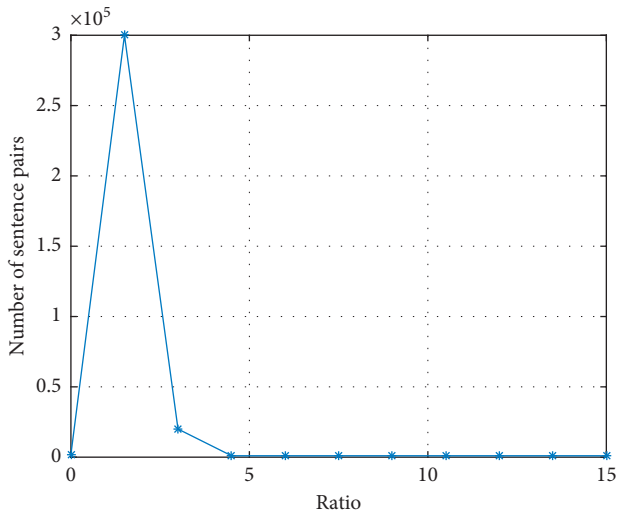


FIGURE 4: Statistics of length-to-length ratios of English-Chinese sentences.

influence of noise and domain differences. We use the training corpus used for statistics. The comparison value of %0.06% falls within the range of 0.4–3.8. Even in the range of 0.5–4.5, there is still 90.11% of the sentence pairs. To this end, we screened the corpus training comparison system for the different ratio ranges of more than 90% of the total corpus and compared the BLEU scores of the systems on the test set. As shown in Table 5, the statistical distribution of the percentage of total corpus pairs in different ratio ranges is listed. The first line represents the corpus used, and the remaining lines represent the number of sentence pairs contained in the different ratio ranges and the percentage of the corpus. Table 5 shows the number of sentence pairs retained and the proportion of the total corpus when the ER filters different pairs of sentences. The higher the score of ER, the better the effect of word alignment, and the more reliable we think the result is. Based on the method of word alignment information screening corpus, we consider two cases: use the filtered sentence to align the retraining words and then get the translation model; directly use the filtered sentence pairs and alignment information to train the translation model. In the first case, the filtered noise information may affect the calculation of the word alignment probability during the iterative process of word alignment. Realignment after filtering out may improve the word alignment quality and improve the translation effect. In the second case, we use ER to retain the word alignment that is considered reliable, so reword alignment or different alignment results may occur due to the change of word alignment probability, and there may be unreliable alignment results in these results. The experimental results are shown in Table 6.

It can be seen from the experimental results that the BLEU scores of each test set are improved in both cases. As far as the overall effect is concerned, it is better not to retrain the word alignment. However, in both cases, the improvement effect on the nist03 and nist05 test sets is not very obvious, the effect of reword alignment on nist03 is slightly better than the latter,

TABLE 5: Scores of 8 length ratios on the development set after filtering.

| $t$   | Ratio range | Number of sentences | Percentage (%) |
|-------|-------------|---------------------|----------------|
| 0.000 | 03–2.0      | 1,911,491           | 100            |
| 0.004 | 0.4–2.0     | 1,876,177           | 98.15          |
| 0.005 | 0.4–1.8     | 1,867,066           | 97.68          |
| 0.010 | 0.4–1.7     | 1,837,419           | 96.12          |
| 0.012 | 0.5–1.7     | 1,814,507           | 94.93          |
| 0.014 | 0.5-i.b     | 1,791,698           | 93.73          |
| 0.020 | 0.5-I.5     | 1,764,233           | 92.30          |
| 0.022 | *           | 1,722,475           | 90.11          |

TABLE 6: BLEU scores based on word alignment filtering (reword alignment).

| ER    | cwmt<br>2011 | nist<br>02 | nist<br>03 | nist<br>04 | nist<br>05 | nist<br>06 | nist<br>08 |
|-------|--------------|------------|------------|------------|------------|------------|------------|
| *     | 22.87        | 30.11      | 28.31      | 29.44      | 27.93      | 24.02      | 18.87      |
| >−0.5 | 23.00        | 30.03      | 27.87      | 29.53      | 27.59      | 24.30      | 19.06      |
| >−0.4 | 22.94        | 30.43      | 28.05      | 29.37      | 27.65      | 24.35      | 19.02      |
| >−0.3 | 23.21        | 30.64      | 28.41      | 29.52      | 27.47      | 24.09      | 19.20      |
| >−0.2 | 23.11        | 30.36      | 28.16      | 29.45      | 27.64      | 24.08      | 19.40      |
| >−0.1 | 23.56        | 29.59      | 28.21      | 29.93      | 27.14      | 24.42      | 19.91      |

and the opposite is on nist45. Use ER to determine whether the word alignment is reliable. When ER is lower than the given threshold, we think that the word alignment result of the sentence pair is not reliable overall. We will filter out the sentence pair, that is, the sentence pair. All alignment information is deleted. In fact, in the word alignment result of the sentence pair, there will be some correct word alignment information; that is, the correct word alignment information is also deleted while deleting the error alignment information. Although the wrong information is not useful for translation tasks, the correct information to be deleted may be helpful for translation tasks. Therefore, there is also the possibility of reducing the BLEU score.

As shown in Figure 5, we find an instance from the filtered sentence pair to illustrate. The thick solid line in the figure is the intersection of two aligned directions, the thin solid line is the correct result of the extended alignment, and the dashed line is the wrong result of the extended alignment. The sentence in Figure 4 is correct for itself, but its alignment information is incorrect; only partial alignment is correct, and its ER value is −0.22; it can be seen that there is some correct word alignment information, and this part of information can be extracted. A rule facilitates translation. Because the ER value is lower than the given threshold, all alignment information of the sentence pair is filtered out, but the correct information is also filtered out, so there is a problem of degraded translation quality.

4.2. Adaptive Experimental Verification of Statistical Machine Translation Based on Massive Corpus. We compared the impact of different searched documents and hidden layer lengths on the accuracy of the translation system. The results are shown in Figure 6.

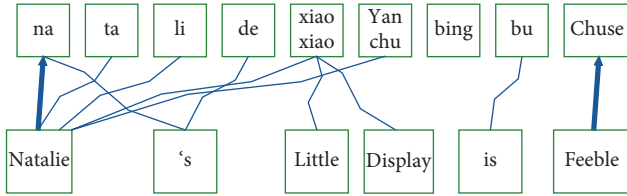


FIGURE 5: Example of incorrect alignment of correct sentence pairs.

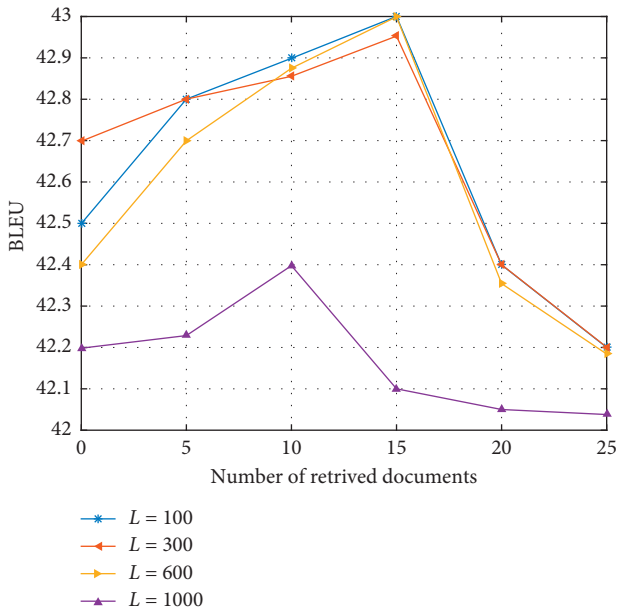


FIGURE 6: The effect of different number of retrieved documents and hidden layer length on machine translation accuracy.

As shown in Figure 6, we found that, for most results, the optimal translation accuracy was obtained when the number of retrieved documents was  $N = 10$ . This result confirms that extending the source language input by means of information retrieval is very helpful for determining topic information. It plays an important role in the selection of translation rules. However, in the experiment, when  $N$  is large, for example,  $N = 50$ , the translation performance is drastically lowered. This is because as the number of retrieved documents increases further, the introduction of topic-independent documents into the neural network will be introduced, and irrelevant documents will bring about unrelated real words, thus affecting the performance of neural network learning.

In Figure 7, it can be seen that when  $L$  is small, the translation system is relatively accurate. In fact, in the case of  $L < 600$ , the difference in translation performance is small. However, when  $L = 1000$ , the translation accuracy is worse than other cases. The main reason is that the amount of parameters in the neural network is so large that it cannot be well studied. We know that when  $L = 1000$ , there are  $100000 \times 1000$  parameters between the linear and nonlinear layers of the network. The current training data size is not enough to support this network parameter level training, so the model is likely to fall into the local Optimal and unacceptable topic representation information.

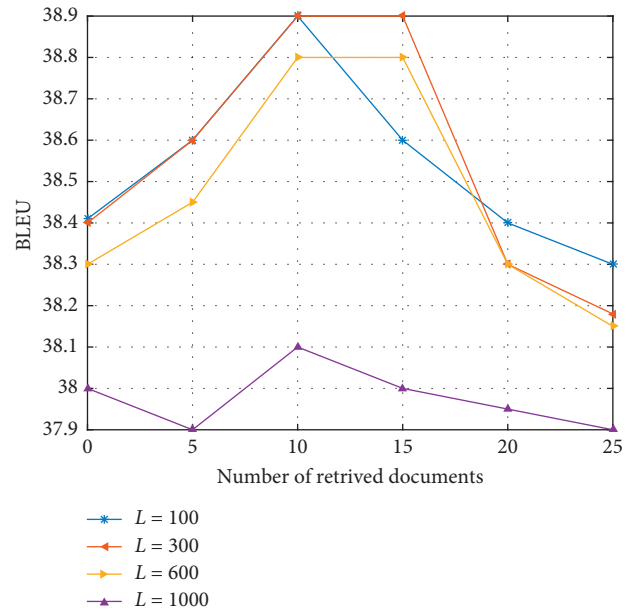


FIGURE 7: The impact of topic-related features on machine translation accuracy.

As shown in Figure 8, we find that the topic similarity feature on the source language side is slightly better for the system than the target language-side similarity feature, and the enhancements that they bring can be accumulated, which means that the neural network based on bilingual data training can help the statistical machine translation system. Translation result candidates perform better disambiguation. Further, based on the similarity feature, the topic sensitivity feature of the translation rule can bring more performance improvement, because the translation rules of specific topics are usually more sensitive, when the similarity is similar. The system tends to choose translation rules for specific topics rather than general translation rules. Finally, we see that our methods perform best when using all topic-related features, with an average of 0.39 BLEU points higher than the LDA-based method.

As shown in Figure 9, we use the method of information retrieval to extend the original input, thus avoiding restrictions on bilingual chapters. We use neural network technology for topic modeling. The algorithm is more practical and has good scalability. Under the deep learning framework, our method directly optimizes the bilingual topic similarity, so that the learned topic representation can be easily integrated into statistical machine translation.

**4.3. Domain Labeling Performance.** The statistics of the data set are used by the domain tagger; 1% is randomly selected as the test data for the domain tagging performance experiment. The training data for training deep learning is selected from the remaining 99% of the size of the data set. At the same time, the domain labeler based on the domain knowledge and the trained deep learning domain labeler are used to label the test data to obtain the category and probability vector, and then the results are simply linearly

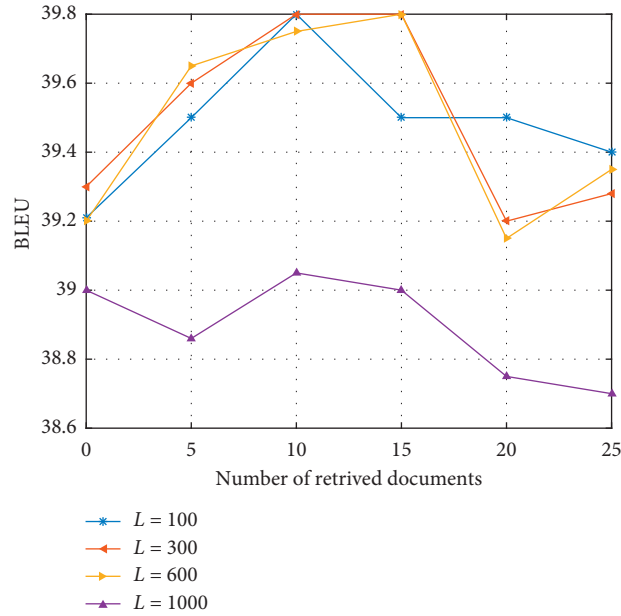


FIGURE 8: The impact of different features on the translation system.

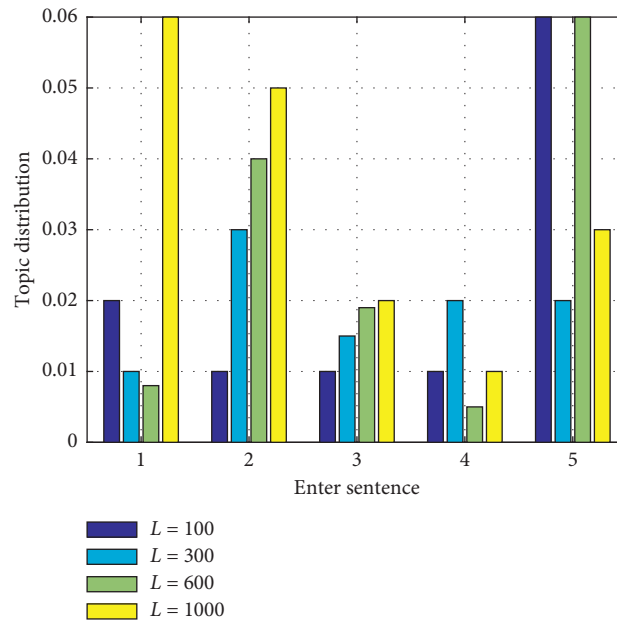


FIGURE 9: Topic distribution information of three sentence candidates.

fused. The category with the highest probability in the probability vector is selected as the final judgment category, and this judgment category is used as the statistical basis for the accuracy rate and the recall rate and the  $F-1$  value. Four subexperiments are performed. The results are shown in Table 7. The results show that using only the domain knowledge tagger will cause misjudgment and omissions due to the lack of self-built domain knowledge base feature words, so the score is not high, but the judgment efficiency is high; using only the deep learning domain tagger requires a lot of the training data belonging to mining tacit knowledge

and continuous features, but training is slow and does not combine prior knowledge; simple linear fusion models linearize the probability vectors of the first two and so on.

Proportional weighting combines explicit and tacit knowledge but this simple fusion is difficult to improve most of the misjudgments and missed judgments and has not been greatly improved; the final neural network deep fusion model is deepened through multilayer neural networks. The integration, giving full play to the advantages of the two, greatly reduced the phenomenon of misjudgment and omission and significantly improved.



TABLE 7: Effect of each labeling method on labeling performance.

|           | Domain knowledge | Deep learning | Simple linear fusion model | Deep fusion model |
|-----------|------------------|---------------|----------------------------|-------------------|
| Accuracy  | 0.85             | 0.9185        | 0.898                      | 0.947             |
| Recall    | 0.76             | 0.95          | 0.886                      | 0.9345            |
| F-1 value | 0.7633           | 0.9531        | 0.8936                     | 0.942             |

## 5. Conclusion

When the training corpus is small, some pairs of training sentences related to the test text may be filtered out, which will affect the quality of the translation. But when the training data is large enough, such problems will hardly occur. In addition to learning the domain model independently for each domain, different domains share the same general model. Through the method of massive corpus, these models can be combined to make the model learning more accurate. The experimental results show that this method can significantly improve the translation accuracy of multiple fields in large-scale machine translation tasks. In addition, the performance of this joint tuning method is better than independent model migration. At the same time, this result can be easily applied to the online translation system, training different models for a pre-determined number of fields, determining the domain according to the input source language text, and selecting the corresponding domain model or general model for translation. The experimental results also show that when there is no such problem, the method of this paper can effectively improve the translation quality of the statistical machine translation system.

The work that needs to be improved in this study is as follows. (1) Consider in the field that the adaptive mechanism is placed in the architecture of human neural machine translation. Calculate different domain vectors, improve the attention mechanism, and make the domain adaptive. (2) How to integrate deep learning methods and prior knowledge to improve the system's performance will be researched in every area of natural language processing in the future. Later, we will try different ways in neural machine translation Chinese and Canadian prior knowledge in the field to improve translation quality for different fields. How to integrate deep learning methods and prior knowledge to improve the performance of the system will be a problem that needs to be studied in the future. After that, we will try to add a priori knowledge in the field of neural machine translation in different ways to improve the translation quality for different fields.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] Y. Liu, C. M. Vong, and P. K. Wong, "Extreme learning machine for huge hypotheses Re-ranking in statistical machine translation," *Cognitive Computation*, vol. 9, no. 2, pp. 285–294, 2017.
- [2] S. Zhu, Y. Yang, M. I. Chenggang, X. Li, and L. Wang, "Corpus selection for Uyghur-Chinese machine translation based on bilingual sentence coverage," *Journal of University of Science & Technology of China*, vol. 47, no. 4, pp. 283–289, 2017.
- [3] K. Kim, E.-J. Park, J.-H. Shin, O.-W. Kwon, and Y.-K. Kim, "Divergence-based fine pruning of phrase-based statistical translation model," *Computer Speech & Language*, vol. 41, pp. 146–160, 2017.
- [4] S. Salami, M. Shamsfard, and S. Khadivi, "Phrase-boundary model for statistical machine translation," *Computer Speech & Language*, vol. 38, pp. 13–27, 2016.
- [5] J. Shang, J. Liu, J. Meng et al., "Automated phrase mining from massive text corpora," *IEEE Transactions on Knowledge & Data Engineering*, vol. 30, no. 10, pp. 1–15, 2018.
- [6] S. Al-Dohuki, Y. Wu, F. Kamw et al., "SemanticTraj: a new approach to interacting with massive taxi trajectories," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 11–20, 2017.
- [7] F. S. Al-Anzi and D. Abuzeina, "Toward an enhanced Arabic text classification using cosine similarity and Latent semantic indexing," *Journal of King Saud University—Computer and Information Sciences*, vol. 29, no. 2, pp. 34–45, 2017.
- [8] M. C. Swain and J. M. Cole, "ChemDataExtractor: a toolkit for automated extraction of chemical information from the scientific literature," *Journal of Chemical Information and Modeling*, vol. 56, no. 10, pp. 1894–1904, 2016.
- [9] Y. Yi, Q. Yao, and H. Qu, "VISTopic: a visual analytics system for making sense of large document collections using hierarchical topic modeling," *Visual Informatics*, vol. 1, no. 1, pp. 45–57, 2017.
- [10] J. P. A. Ioannidis, "The reproducibility wars: successful, unsuccessful, uninterpretable, exact, conceptual, triangulated, contested replication," *Clinical Chemistry*, vol. 63, no. 5, pp. 943–945, 2017.
- [11] M. Rahnemoonfar, G. C. Fox, M. Yari, and J. Paden, "Automatic ice surface and bottom boundaries estimation in radar imagery based on level-set approach," *IEEE Transactions on Geoscience & Remote Sensing*, vol. 55, no. 9, pp. 1–8, 2017.
- [12] J. Thomas and A. Zaytseva, "Mapping complexity/human knowledge as a complex adaptive system," *Complexity*, vol. 21, no. S2, pp. 207–234, 2016.
- [13] F. Baertling, B. Alhaddad, A. Seibt et al., "Neonatal encephalocardiomyopathy caused by mutations in VARS2," *Metabolic Brain Disease*, vol. 32, no. 1, pp. 1–4, 2016.
- [14] M. A. Raquet, G. J. Measey, and J. M. Exbrayat, "Annual variation of ovarian structures of Boulengerula taitana (Loveridge 1935), a Kenyan caecilian," *African Journal of Herpetology*, vol. 64, no. 2, pp. 116–134, 2016.
- [15] C. Hedberggoldfors, N. Darin, and A. Oldfors, "Muscle pathology in Vici syndrome—a case study with a novel mutation in EPG5 and a summary of the literature," *Neuromuscular Disorders*, vol. 27, no. 8, pp. 840–858, 2017.
- [16] E. X. Fang, M. D. Li, M. I. Jordan, and H. Liu, "Mining massive amounts of genomic data: a semiparametric topic modeling

- approach,” *Journal of the American Statistical Association*, vol. 112, no. 519, pp. 1–15, 2017.
- [17] M. Knight, “The accused is entering the courtroom: the live-tweeting of a murder trial,” *Journal of Media Practice*, vol. 18, no. 2-3, pp. 1–26, 2017.
- [18] K. Hörtnagel, I. Krägeloh-Mann, A. Bornemann et al., “The second report of a new hypomyelinating disease due to a defect in the VPS11 gene discloses a massive lysosomal involvement,” *Journal of Inherited Metabolic Disease*, vol. 39, no. 6, pp. 849–857, 2016.
- [19] A. Takagi, H. Ozawa, M. Oki, H. Yanagi, K. Nabeshima, and N. Nakamura, “*Helicobacter pylori*-negative advanced gastric cancer with massive eosinophilia,” *Internal Medicine*, vol. 57, no. 12, pp. 1715–1718, 2018.
- [20] W. Kojima and K. Hayashi, “Changes in the axo-glial junctions of the optic nerves of cuprizone-treated mice,” *Histochemistry & Cell Biology*, vol. 149, no. 5, pp. 1–8, 2018.
- [21] M. Kala, M. V. Shaikh, and M. Nivsarkar, “Equilibrium between anti-oxidants and reactive oxygen species: a requisite for oocyte development and maturation,” *Reproductive Medicine and Biology*, vol. 16, no. 1, pp. 28–35, 2017.
- [22] M. Betanzos, M. R. Costa-jussà, and L. Belanche, “Tradares: a tool for the automatic evaluation of human translation quality within an MOOC environment,” *Applied Artificial Intelligence*, vol. 31, no. 1, pp. 1–10, 2017.

## Research Article

# A Multiscale and High-Precision LSTM-GASVR Short-Term Traffic Flow Prediction Model

Jingmei Zhou <sup>1</sup>, Hui Chang <sup>2</sup>, Xin Cheng <sup>2</sup> and Xiangmo Zhao <sup>2</sup>

<sup>1</sup>School of Electronic and Control Engineering, Chang'an University, Xi'an 710064, China

<sup>2</sup>School of Information Engineering, Chang'an University, Xi'an 710064, China

Correspondence should be addressed to Jingmei Zhou; [jmzhou@chd.edu.cn](mailto:jmzhou@chd.edu.cn), Hui Chang; [480796959@qq.com](mailto:480796959@qq.com), and Xin Cheng; [xincheng@chd.edu.cn](mailto:xincheng@chd.edu.cn)

Received 28 March 2020; Revised 23 April 2020; Accepted 28 April 2020; Published 17 June 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Jingmei Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Short-term traffic flow has the characteristics of complex, changeable, strong timeliness, and so on. So the traditional prediction algorithm is difficult to meet its high real-time and accuracy requirements. In this paper, a multiscale and high-precision LSTM-GASVR short-term traffic flow prediction algorithm is proposed. This method uses 15 min traffic flow data of the first 16 sections as input and completes the data preprocessing operation through reconstruction, normalization, and rising dimension by working day factor; establishing the prediction model based on the long- and short-term memory network (LSTM) and inverse normalization; and proposing the GA-SVR model to optimize the prediction results, so as to realize the real-time high-precision prediction of traffic flow. The prediction experiment is carried out according to the charge data of a toll station in Xi'an, Shaanxi Province, from May 2018 to May 2019. The comparison and analysis of various algorithms show that the prediction algorithm proposed in this paper is 20% higher than the LSTM, GRU, CNN, SAE, ARIMA, and SVR, and the  $R^2$  can reach 0.982, the *explanatory variance* is 0.982, and the MAPE is 0.118. The proposed traffic flow prediction algorithm provides strong support for traffic managers to judge the state of the road network to control traffic and guide traffic flow.

## 1. Introduction

High-precision and short-term traffic flow forecast shows the future trend of traffic development [1]. For intelligent traffic management, road network planning provides future traffic flow data [2]. It is not only helpful to alleviate traffic congestion, but also important for autonomous vehicles [3]. Traffic flow prediction refers to forecast traffic flow in the next period based on several historical traffic flow data [4]. A prediction one to several hours ahead of schedule is often called a short-term prediction [5, 6]. As a hot theme in Intelligent Transportation [7–9], there were lots of relating research studies about short-term traffic flow prediction. Wang et al. [10] proposed a hybrid short-term traffic speed prediction framework based on empirical mode by decomposition (EMD) and autoregressive integrated moving average (ARIMA) and achieved short-term traffic flow prediction of expressway in different scenarios, but there are

great differences in the prediction effect for different types of vehicles. Vasantha Kumar and Vanajakshi [11] developed a SARIMA short-term traffic flow prediction model under constrained data. The model needed high stable data and was weak at generalization. A two-step prediction method based on stochastic differential equation (SDE) was implemented in reference [12], which improves the prediction accuracy of a periodic data, but neglects the prediction effect of periodic traffic flow. Salamanis et al. [13] studied a density-based clustering method which can accurately predict traffic flow under normal and abnormal conditions. The prediction effect is good, but the time complexity is high and the feasibility is slightly lower. Neuhold et al. [14] established a model to predict traffic flow and an algorithm to optimize lane allocation in front of the toll plaza in Austria. The prediction model and optimization algorithm are not site specific and can be applied to different toll plaza or expressway bottlenecks (such as road engineering, transit, and

highway intersections). But for the data that are one-day or one-hour interval traffic flow, this model cannot fully meet the intelligent traffic management needs. A prediction model based on deep recurrent neural network (RNN) was designed in reference [15]. While the prediction effect is better, there are great differences in the prediction effect of a periodic data. Qu et al. [16] used historical traffic flow data and environmental factor data to predict all-day traffic flow based on deep neural networks, using multilayer supervised learning algorithm to train predictors, mining the potential relationship between traffic flow data and key contextual factors, an intermittent training method for reducing training time. But the prediction effect for smaller traffic flow data is poor. In [17], based on the BP neural network to predict the expressway traffic flow, the incentive function of the model is improved, and to some extent, the prediction effect is improved, but it is not suitable for all highways, and the limitation is great. In reference [18], the traffic flow information acquisition technology and combination model of video image analysis were applied to the traffic flow prediction to reduce the error and time for establishment. But the prediction effect is affected by the accuracy of data acquisition and fluctuates greatly. Luo et al. [7] implemented a CNN-SVR (convolutional neural networks-support vector regression) short-term traffic flow prediction model. The Adam optimization algorithm was used to ensure the completeness of the time-space flow characteristics, which reduces the interference of external factors and can effectively predict the traffic flow. However, because the data are relatively single, the generalization ability of the model is not strong. In reference [19], a stationary short-time traffic flow prediction method was proposed. The support vector machine predicts the traffic flow data after stationary, which solves the influence of the asymmetric distribution of the data on the prediction effect, but caused a bad prediction effect of the extreme point. Wang et al. [20] proposed an improved BP neural network model based on thought evolution algorithm. The accuracy of the prediction model was improved by phase space reconstruction of traffic flow time series using chaotic theory. However, because the model uses less data, lacks credibility, and did not consider the prediction effect of special time periods such as holidays, it is not universal. A traffic flow prediction model based on convolutional neural network is implemented in reference [21], and the paper considered the spatial-temporal correlation of regional traffic flow, which improves the prediction accuracy and stability of the model. Only taxi traffic flow data were studied, which do not fully reflect the road state. Previous neural network models used values from previous days or weeks as inputs to predict future traffic flow data [22, 23], and the real time is poor, which exists the problem of gradient vanishing. To address this, LSTM (Long Short-Term Memory) neural network was used to predict traffic flow [24–27], but the influence of commuting traffic flow on urban traffic state was not considered.

Tidal traffic flow phenomenon is one of the important factors causing congestion in toll stations. Accurate and real-time prediction of short-term traffic flow provides strong support for traffic management departments to guide traffic

flow. And it plays an important role in the information communication system [28]. This paper presents a multi-scale and high-precision LSTM-GASVR (Long Short-Term Memory Genetic Algorithm Support Vector Regression) short-term traffic flow prediction algorithm. Data preprocessing was utilized to filter, normalize, and reconstruct the data and improve the data dimension by workday flags, which effectively improves the convergence speed and prediction accuracy of the model; building prediction model based on the output LSTM, Long Short-Term Memory network, and GA-SVR model was proposed to optimize the prediction results to achieve real-time and high-precision prediction of short-term traffic flow. The proposed traffic flow prediction algorithm improves the prediction accuracy and stability and improves the generality and feasibility of the traffic flow prediction model.

## 2. Short Traffic Flow Data Preprocessing

In order to obtain the appropriate amount of data and the required data format, traffic flow data need to be pre-processed by data merging, expanding the amount of data, and data resampling to obtain time traffic volume. The preprocessing of short-term traffic flow data is divided into the following steps, as shown in Figure 1:

- (i) Data normalization makes the data model training have better convergence
- (ii) Considering the occasional and violent weekend and holiday tidal traffic flow, it is necessary to enhance the data dimension
- (iii) Data reconstruction aims to obtain a specific data format

Step 1: extract the data and merge the data for  $n$  consecutive months as needed to expand the dataset.

Step 2: resample the data as needed. Table 1 shows the charge data of Shaanxi Provincial Toll Station in August 2018; “STARTTIME” is the computer boot time; the “ENTRYLANE” is the entrance lane number; the “ENTRYTIME” indicates the vehicle entry time; and the “VEHICLETYPE” is the vehicle type; 0, 1, and 2, respectively, indicate the uncertain vehicle type, passenger car, and freight car; “VEHICLELICENSE” represents the license plate number. To extract the “ENTRYTIME” column of Table 1, we regard one line of data as a vehicle entering the station. These data are resampled according to 15 min, and the data shown in Table 2 are obtained. The “Value” represents the traffic volume within 15 min with the starting time of “Date.”

Step 3: data normalization. For a better convergence of the prediction results and the better results, the data are normalized by the maximum and minimum value standardization, assuming that the  $x$  is the traffic volume of a certain period of time, and it is normalized as follows:

$$x' = \frac{x - \min}{\max - \min}, \quad (1)$$

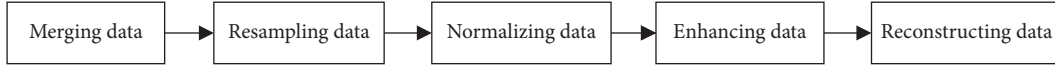


FIGURE 1: Data preprocessing.

TABLE 1: Partial data of a toll station in Shaanxi Province.

| STARTTIME          | ... | ENTRYLANE | ENTRYTIME         | VECHILETYPE | ... | VECHILELICENSE | ... |
|--------------------|-----|-----------|-------------------|-------------|-----|----------------|-----|
| 2018/7/29 19:22:51 | ... | 12        | 2018/8/1 14:05:05 | 1           | ... | (B)S ADT702    | ... |
| 2018/8/1 11:55:01  | ... | 7         | 2018/8/1 14:05:05 | 1           | ... | (B)E FLG085    | ... |
| 2018/7/29 19:22:51 | ... | 12        | 2018/8/1 14:05:02 | 1           | ... | (B)S A53H75    | ... |
| ...                | ..  | ...       | ...               | ...         | ..  | ...            | ... |

TABLE 2: Partial traffic data of 15 min at Qujiang Toll Station, Shaanxi Province.

| Date          | Value |
|---------------|-------|
| 2018/8/1 8:00 | 186   |
| 2018/8/1 8:15 | 239   |
| 2018/8/1 8:30 | 300   |
| ...           | ...   |

TABLE 3: Format of traffic flow forecast training data for  $m$  period before the day.

| Sample no.                          | 0           | 1           | ... | $n$         |
|-------------------------------------|-------------|-------------|-----|-------------|
| Period 0 traffic volume (input)     | $v_0^0$     | $v_0^1$     | ... | $v_0^n$     |
| Period 1 traffic volume (input)     | $v_1^0$     | $v_1^1$     | ... | $v_1^n$     |
| ...                                 | ...         | ...         | ... | ...         |
| Period $m-1$ traffic volume (input) | $v_{m-1}^0$ | $v_{m-1}^1$ | ... | $v_{m-1}^n$ |
| Working day factors                 | $h_b^0$     | $h_b^1$     | ... | $h_b^n$     |
| Next period traffic volume (output) | $v_m^0$     | $v_m^1$     | ... | $v_m^n$     |

$x'$  is the values after  $x$  normalization fall, from 0 to 1, and the min and max in equation (1) are the minimum and maximum values in the dataset.

Step 4: dimension promotion. The traffic demand caused by commuter's commute is an important part of urban road traffic volume. Because of the urban planning problems, the traffic flow of commuters has an obvious influence on urban traffic state. The working day factors are divided into three categories according to "week," "weekend," and "holiday."

Step 5: data reconstruction. The data are reconstructed as needed, the data are divided into  $m+2$  row and  $n+1$  column, the first  $m+1$  row of each column is input, and the last row is compared with the predicted value as the real value. The final data structure is shown in Table 3.  $v_i^j$  represents traffic volume in sample  $i$  at period  $j$ .  $h_b^j$  represents the working day factor in sample  $j$ .  $h_b$  represents the working day factor,  $b$  is 1, 2, or 3,  $h_1 = 0.1$ ,  $h_2 = 0.5$ , and  $h_3 = 0.9$ .

### 3. Analysis of Time Series Predictive Assessment Indicators

The prediction results of time series need to evaluate each prediction model through indicators and adjust the parameters, so as to determine the high-precision short-term traffic flow prediction model. Equations (2)–(9) are evaluation indicators of time series prediction [2].  $f_i$  represents the predicted value of sample  $i$ ,  $y_i$  represents the actual value of sample  $i$ ,  $\bar{y}$  represents the average value of the real data, and  $n$  is the sample size.

① MAE (mean absolute error):

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |f_i - y_i|. \quad (2)$$

② MSE (mean square error is one of the most commonly used performance measures for regression tasks):

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (f_i - y_i)^2. \quad (3)$$

③ RMSE (root mean squared error):

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - y_i)^2}. \quad (4)$$

④ MAPE (mean absolute percentage error) often used to measure prediction accuracy:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - f_i}{y_i} \right|. \quad (5)$$

⑤ Explanatory variance:

The explanatory variance represents the extent to which the regression model explains the variation of the variance of the dependent variable (i.e., fitting effect of the regression model to the true value), which is a common evaluation index for the regression model:

$$\text{Explained\_variance}(y, f) = 1 - \frac{\text{Var}\{y - f\}}{\text{Var}\{y\}}. \quad (6)$$

The explanatory variance is between  $[0, 1]$ , and the closer to 1 indicates the better the prediction effect.

⑥  $R^2$  ( $R$ -squared):

$R^2$  is the ratio of the sum of squared residuals to the sum of the total deviation squares, indicating the degree to which the regression equation can explain the variation of the dependent variable, and  $R^2$  is the fitting effect of the regression equation on the true value.

The sum of squared residuals:

$$SS_{\text{res}} = \sum_{i=1}^n (f_i - y_i)^2. \quad (7)$$

The sum of total deviation squares:

$$SS_{\text{tot}} = \sum_{i=1}^n (y_i - \bar{y})^2, \quad (8)$$

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} = 1 - \frac{\sum_{i=1}^n (f_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad (9)$$

The value of  $R^2$  is between 0 and 1. The closer  $R^2$  is to 1, the more accurate the model and the better the regression prediction effect. It is generally considered that the model fit is higher when  $R^2$  exceeds 0.8.

Because ①②③ use the mean error, and the mean error is more sensitive to outliers, if there is a large difference between a certain regression value fitted by the regressor and the true value, it will lead to a large average error, which will have a great influence on the final evaluation value, that is, the mean value is not robust. According to Table 4, it is found that the value of ①②③ increases with the increase of  $y$ , so the reliability is low. By comparing and analyzing the curve fitting effect in Figures 2–13, in combination with the value of evaluation indicators in Tables 4–6, it is found that the evaluation value of ④⑤⑥ is more consistent with the curve fitting effect. By analyzing and comparing the above evaluation indicators, take ④⑤⑥ as the main index to evaluate traffic flow prediction and ①③ as the auxiliary index.

## 4. Short-Term Traffic Flow Prediction Algorithm Based on LSTM-GASVR

*4.1. Optimize the Model Based on GA-SVR.* Genetic algorithm (GA) is used to optimize the parameters  $(C, \varepsilon, r)$  in the SVR model, where the  $C$  is the penalty coefficient,  $r$  is the kernel function coefficient, and  $\varepsilon$  is the insensitive coefficient, which has the following main steps:

- (1) Constructing the chromosome assemblage  $(C, \varepsilon, r)$  and formulating the fitness calculation function of genetic algorithm
- (2) Determining the parameters of selection, crossover, mutation, and so on in the GA, and setting the iterative termination condition of the algorithm
- (3) Initializing the GA and generating the initialization population
- (4) Calculating individual fitness in chromosome populations
- (5) Generation of next-generation chromosomes through selection, crossover, variation, etc.

TABLE 4: Comparison of short-term traffic flow prediction effect at different time intervals.

| Time intervals | $R^2$ | Explanatory variance | MAPE  | MAE    | RMSE   |
|----------------|-------|----------------------|-------|--------|--------|
| 5              | 0.980 | 0.980                | 0.232 | 7.705  | 11.532 |
| 10             | 0.979 | 0.981                | 0.186 | 13.408 | 19.777 |
| 15             | 0.982 | 0.982                | 0.160 | 21.406 | 27.960 |
| 20             | 0.980 | 0.980                | 0.156 | 28.095 | 40.394 |
| 25             | 0.974 | 0.977                | 0.219 | 36.871 | 54.758 |
| 30             | 0.971 | 0.972                | 0.168 | 49.940 | 69.037 |

- (6) If the iterative termination condition of the algorithm is satisfied or not, jump to (4).
- (7) Termination of the iteration to determine the optimal parameters  $(C, \varepsilon, r)$ . Figure 14 shows a flowchart of GA solving SVR optimal parameters  $(C, \varepsilon, r)$ .

*4.2. Chromosome Coding.* Genetic method is based on chromosome coding, selection, crossover, and mutation, and other operations of the algorithm effectively depend on the chromosome coding method to some extent. Considering that the SVR model parameter selection itself is a constrained problem, we code the SVR model parameter by real number. That is the gene of each chromosome consists of three decimal float.

*4.3. Population Initialization.* Population number is an important parameter that affects the efficiency and convergence of the algorithm. In this paper, the population size  $\text{pop\_size}$  is set to 50, and the population initialization state is evenly distributed in the solution space. And  $C \in [275, 285]$ ,  $\varepsilon \in [0.4, 0.6]$ , and  $\sigma \in [0.001, 0.003]$ .

*4.4. Fitness Function Selection.* GA is used to find the best parameters of the SVR model, so the fitness function is chosen as average relative error percentage between the prediction results and the actual value, and the fitness function can be designed as follows:

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \frac{|y_i - f(x_i)|}{y_i}. \quad (10)$$

Among them, the MAPE is the average relative error percentage between the prediction results and the actual value of the algorithm.  $t_0$  is the initial time of training data,  $y_i$  represents the actual driving speed of the road at the time  $t_0 + i \times 15$  min, and the prediction result of the time  $t_0 + i \times 15$  min is  $f(x_i)$ .

*4.5. Selection Operations.* The selection operation used in this paper is the combination of roulette and elite strategy. The key of the roulette selection method is to produce offspring population by calculating the probability of each individual appearing in the offspring, that is, the greater the probability of individual selection when the adaptation value is higher. The probability of individual appearance is shown in Table 7. Roulette selection algorithm refers to the

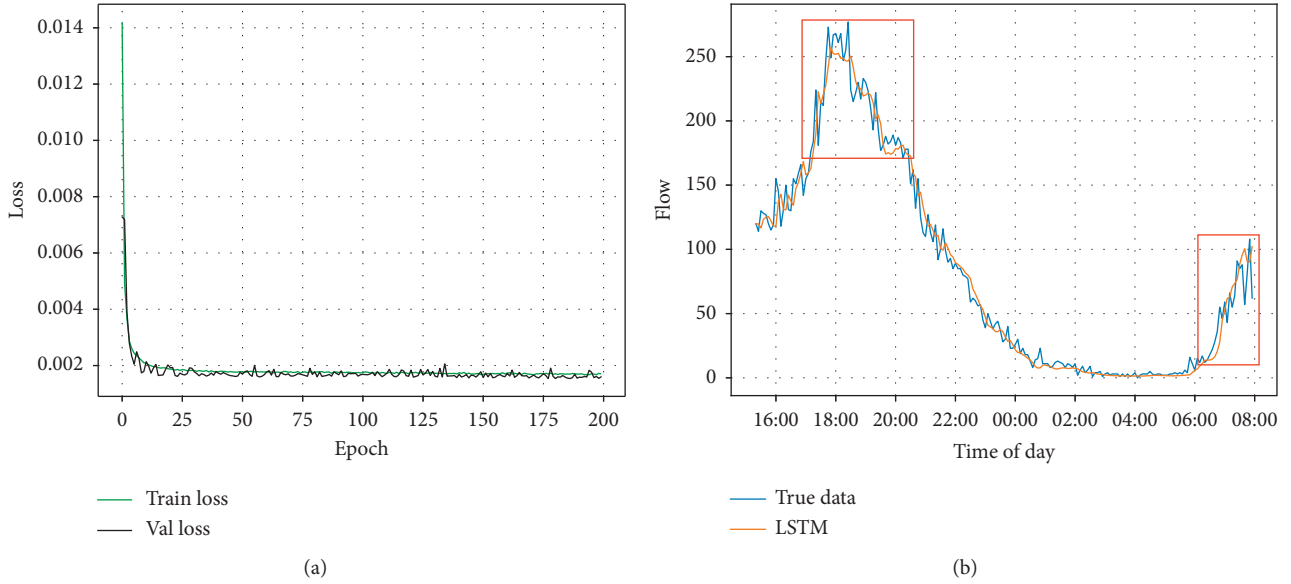


FIGURE 2: 5 min LSTM training and prediction. (a) 5 min LSTM loss function and (b) 5 min LSTM prediction results.

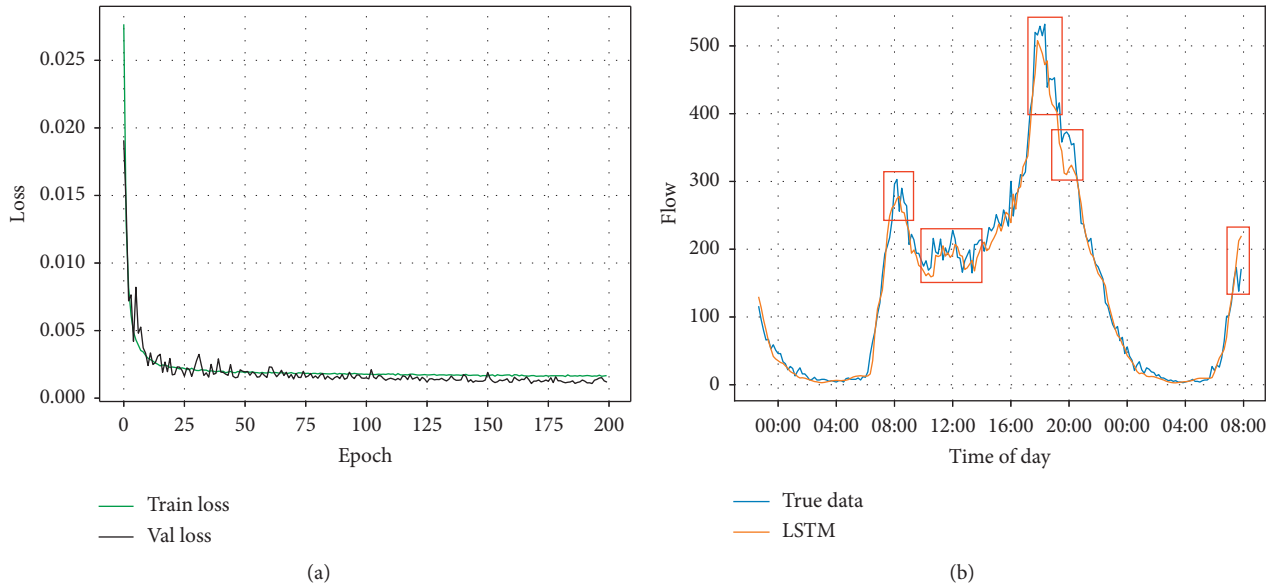


FIGURE 3: 10 min LSTM training and prediction. (a) 10 min LSTM loss function and (b) 10 min LSTM prediction results.

generation of random numbers in the interval  $[0, 1)$ , which falls within a certain probability interval shown in Table 7, which corresponds to individuals inherited to the next generation. For example, the random number generated is 0.3,  $0.3 \in [0.2, 0.45)$ , corresponding to individual 2, so individual 2 is inherited to the next generation. The disadvantage of this method is that it is easy to fall into local optimum when the range of adaptive value interval is small.

The elite strategy is to keep the better individuals in the last generation population and increases the number of the better individuals so as to guarantee the global optimum, but it is easy to fall into the local optimum when the proportion of the better individuals is large. Elite strategies retain better

individuals by copy better individuals into the next generation or cross and mutation better individuals in the population, and we select the latter method.

Therefore, the paper adopts the combination of roulette and elite strategy to keep the better population number in elite strategy proportionally dynamic, so as to get the solution of the population number, as shown in the following equation:

$$\text{GoodPop\_size} = \min\left(\text{Pop\_size}, \frac{k \times \text{current\_Gen}}{\text{Max\_Gen}}\right). \quad (11)$$

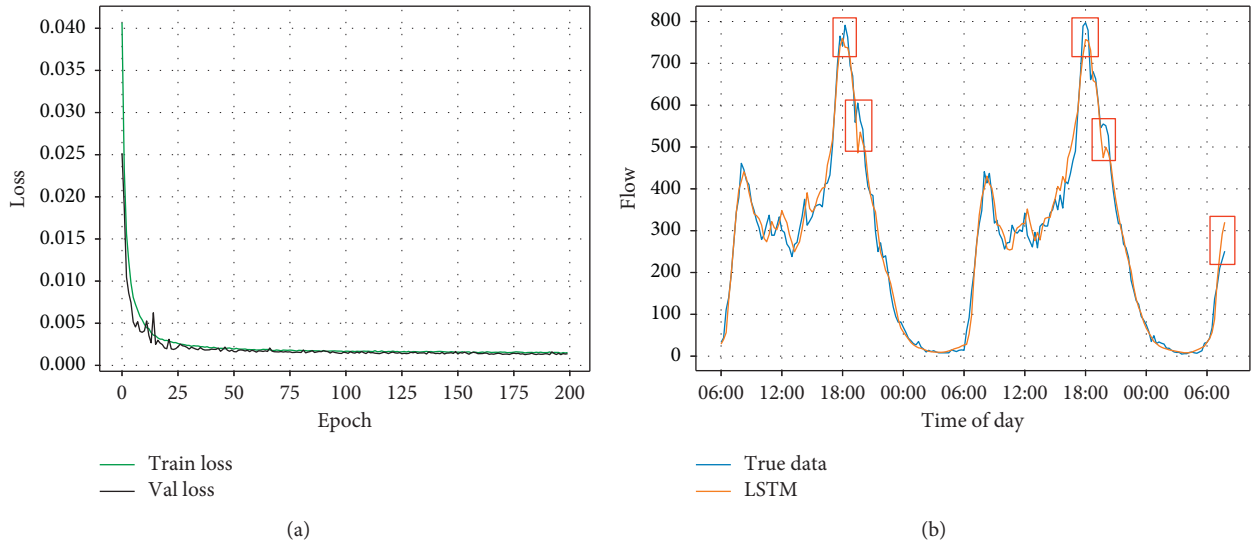


FIGURE 4: 15 min LSTM training and prediction. (a)15 min LSTM loss function and (b) 15 min LSTM prediction results.

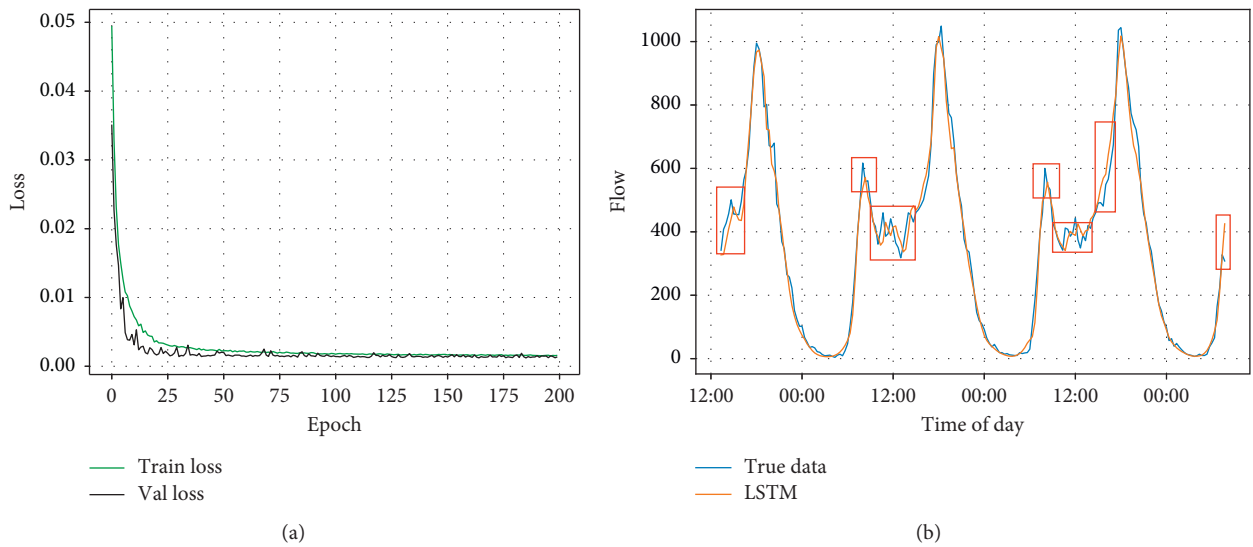


FIGURE 5: 20 min LSTM training and prediction. (a) 20 min LSTM loss function and (b) 20 min LSTM prediction results.

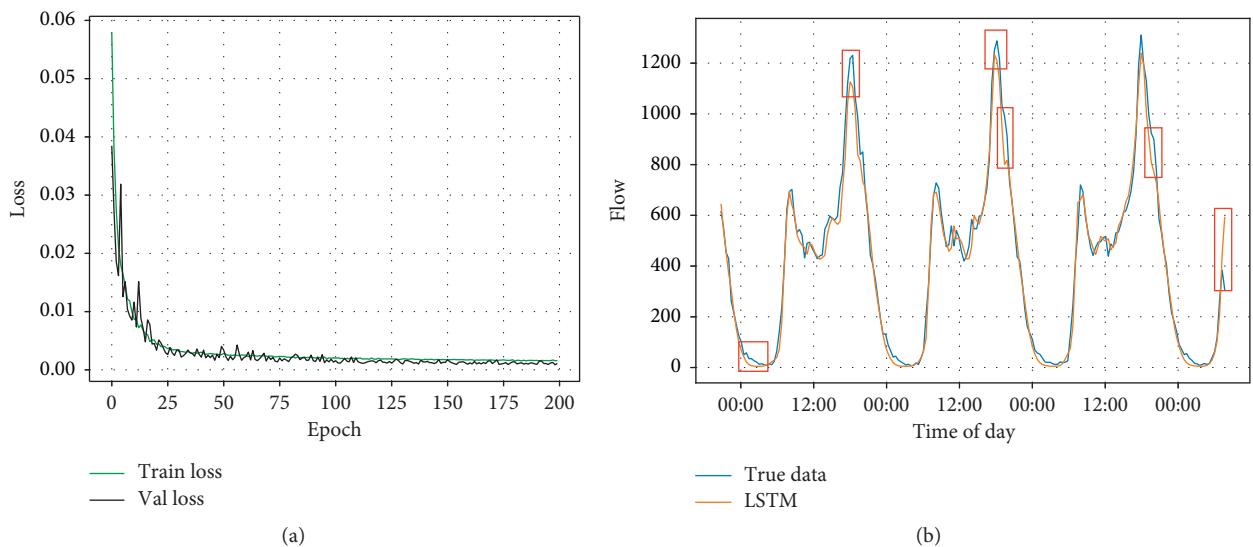


FIGURE 6: 25 min LSTM training and prediction. (a) 25 min LSTM loss function and (b) 25 min LSTM prediction results.



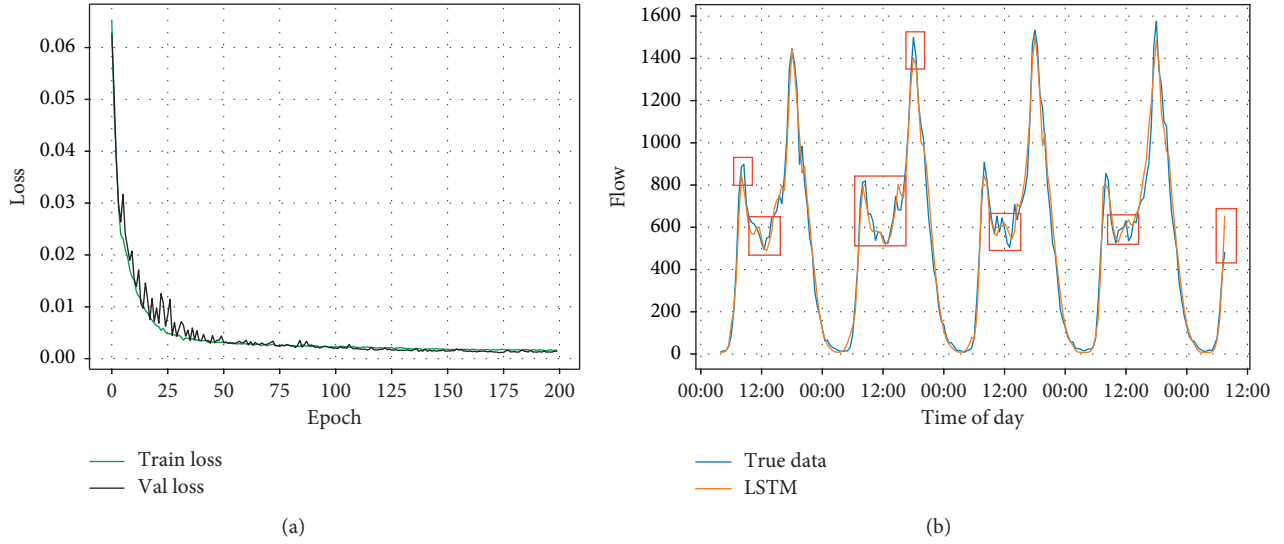


FIGURE 7: 30 min LSTM training and prediction. (a) 30 min LSTM loss function and (b) 30 min LSTM prediction results.

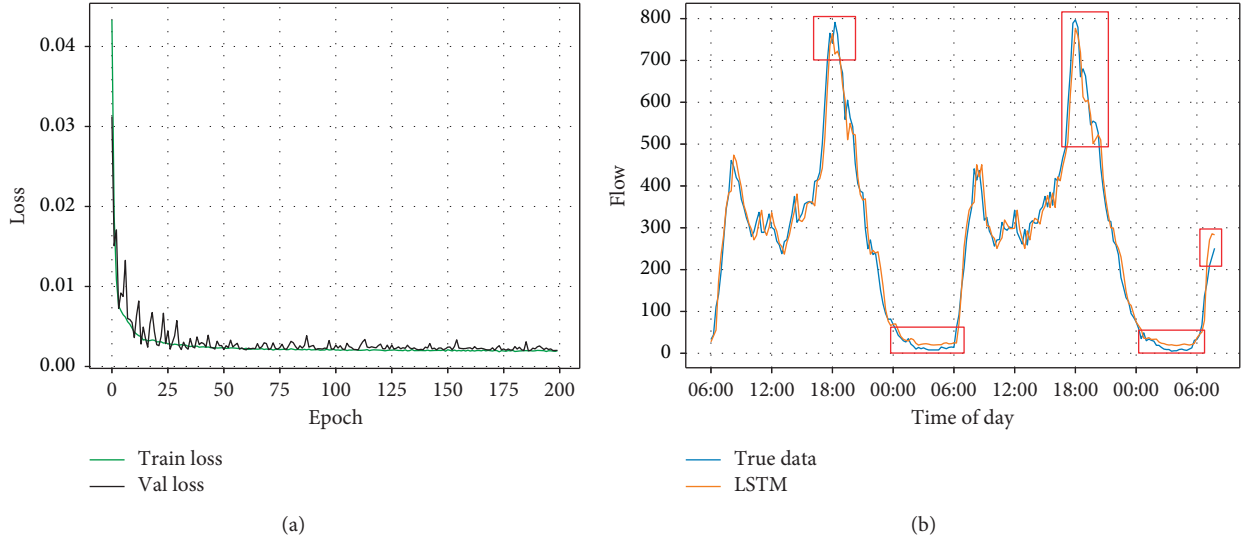


FIGURE 8: 4 steps LSTM training and prediction. (a) 4 steps LSTM loss function and (b) 4 steps LSTM prediction results.

The GoodPop\_size in equation (11) is to retain the number of better individuals, Pop\_size represent the number of populations,  $k$  is the sum of the parent and the offspring at each run time of the algorithm, current\_Gen is the number of runs of the current algorithm, and Max\_Gen is the maximum number of iterations set in advance for the algorithm. Equation (11) provides a corresponding guarantee for the efficiency and optimization performance.

**4.6. Genetic Operator Design.** Genetic operator mainly includes cross operator and mutation operator. Cross operation is one of the main operations of offspring innovation in GA. By simulating the cross inheritance of chromosome weight, the individual genes are exchanged according to certain probability. The mutation operation is to simulate the process of chromosome compilation and recombination

to change some genes in individuals according to certain probability. Cross operator and mutation operator are a way to produce innovation.

**4.6.1. Cross Operator.** During the evolution of natural organisms, the recombination of biological genetic genes is very important, and cross operators cannot be replaced in GA. The arithmetic cross operator is selected in the paper, as shown in equation (12).  $a$  in equation (12) represents the random number between 0 and 1 of the values of cross probability.  $P'_i$  and  $P'_{i+1}$  represent gene value  $P_i$  and  $P_{i+1}$  after cross operator, and  $i$  and  $i+1$  are gene positions:

$$P'_i = aP_i + (1-a)P_{i+1}, \quad P'_{i+1} = (1-a)P_{i+1} + aP_i. \quad (12)$$

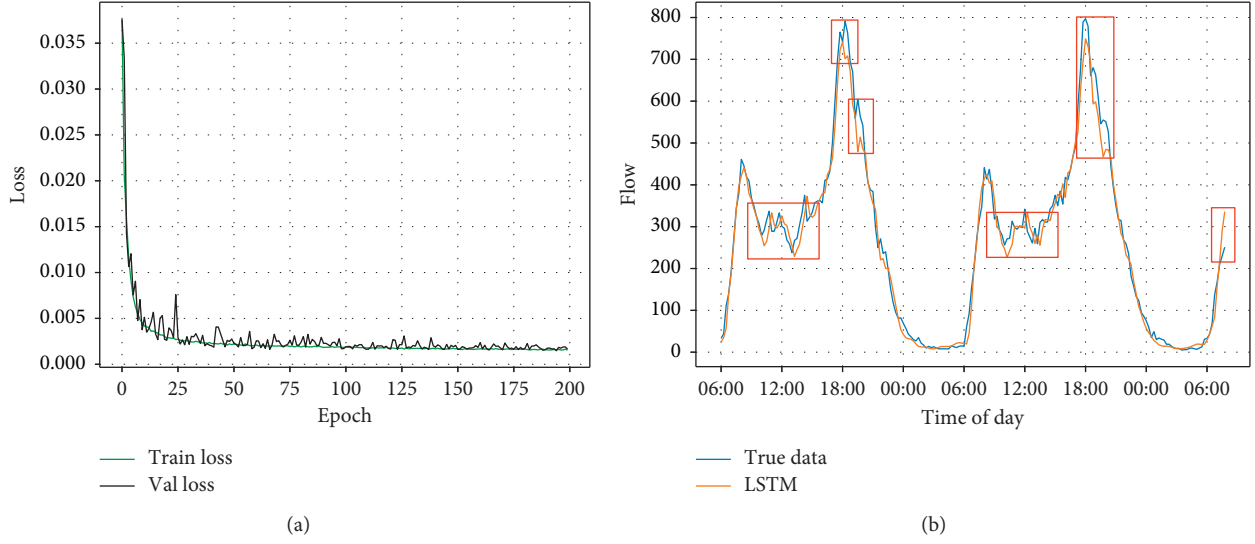


FIGURE 9: 8 steps LSTM training and prediction. (a) 8 steps LSTM loss function and (b) 8 steps LSTM prediction results.

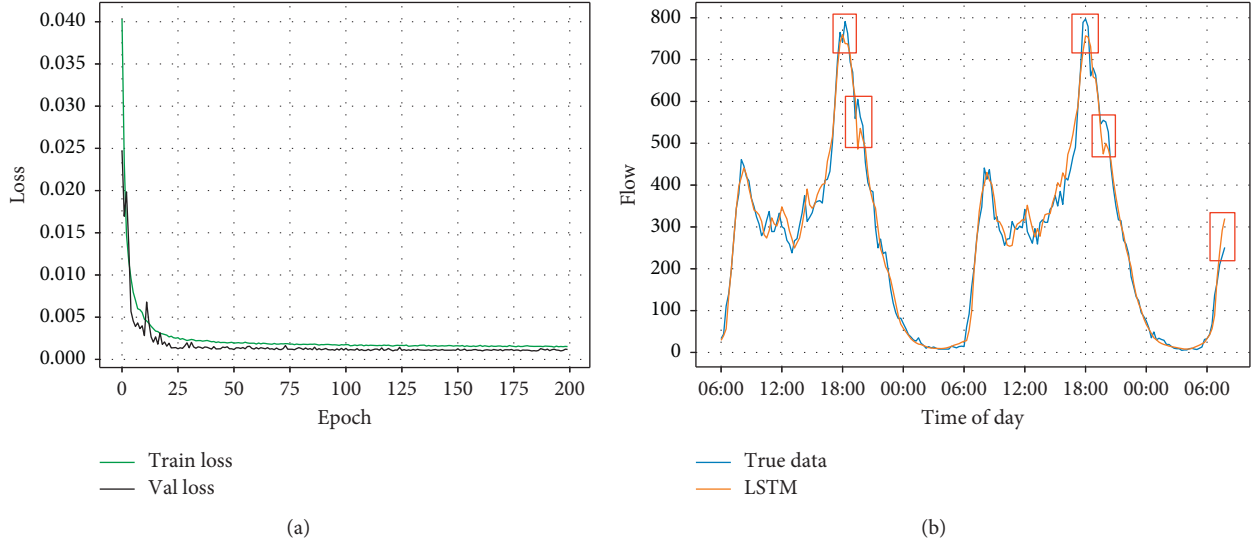


FIGURE 10: 12 steps LSTM training and prediction. (a) 12 steps LSTM loss function and (b) 12 steps LSTM prediction results.

**4.6.2. Mutation Operator.** Mutation operator is the operation of changing individual gene value  $x_j$  in chromosome population, as shown in equation (13). A random function  $\text{Random}(a_i, b_i)$  is used to generate a random number, and  $a_i$  and  $b_i$  are the upper and lower limits of the gene. And a gene position  $j$  in the chromosome is set to the value generated by the random function:

$$x_j = \begin{cases} \text{Random}(a_i, b_i), & i = j, \\ x_j, & i \neq j. \end{cases} \quad (13)$$

**4.7. Stop Sign.** The stop flag is the maximum number of iterations set for the experiment, and the value is set to 200 in the experiment.

During the GA training as shown in Figure 15, the blue boxes indicates where the optimal value goes up and the

optimal value increased to 0.86 after 30 iterations. The red curve in the figure represents the accuracy of the optimal parameter SVR of the prediction model. Because of the elite retention strategy, the green curve gradually approaches red, that is, the average accuracy tends to the optimal value. The SVR model parameters of the final GA tuning parameter  $(C, \epsilon, r)$  are  $(276.7, 0.05998, 0.001595)$ .

**4.8. LSTM Prediction of Short-Term Traffic Flow Based on GA-SVR Optimization.** Based on the research and analysis of various algorithms, it is found that the LSTM model prediction evaluation index  $R^2$  is higher, that is, the fitting degree between the predicted data and the real traffic flow data is higher, which is more consistent with the real data trend, and the prediction results are more reliable. GA-SVR prediction model evaluation index  $MAPE$  is smaller, the

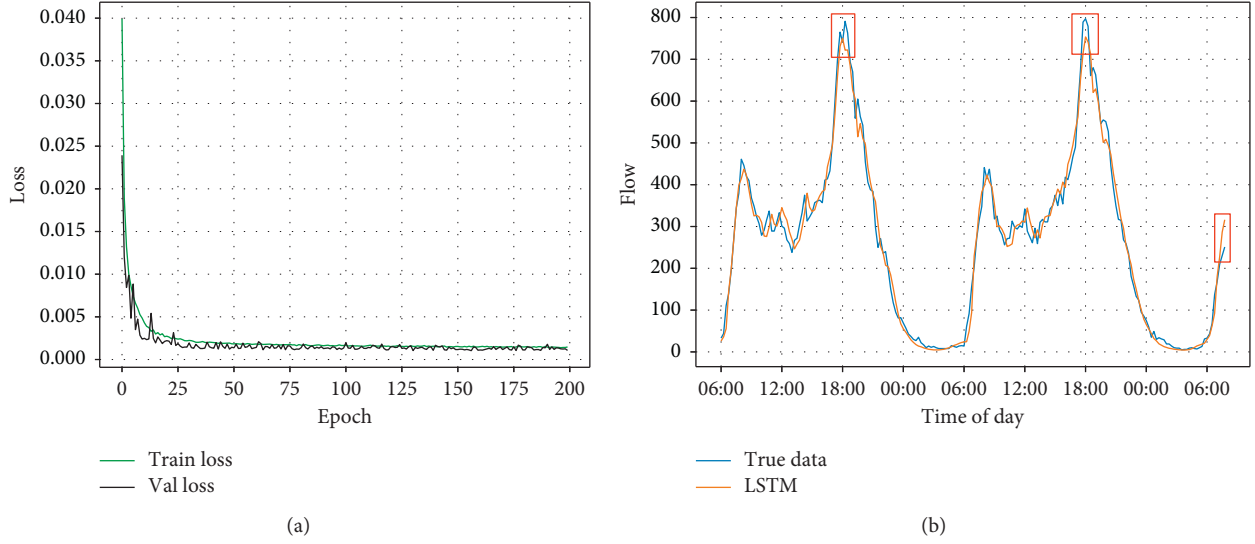


FIGURE 11: 16 steps LSTM training and prediction. (a) 16 steps LSTM loss function and (b) 16 steps LSTM prediction results.

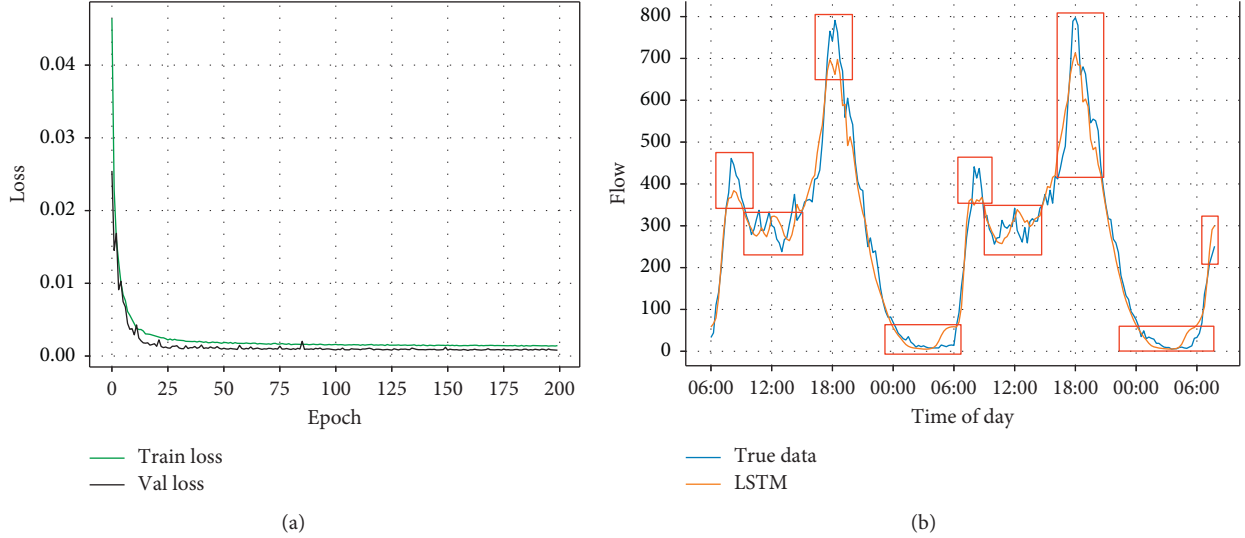


FIGURE 12: 20 steps LSTM training and prediction. (a) 20 steps LSTM loss function and (b) 20 steps LSTM prediction results.

model prediction value is closer to the real traffic flow data numerically, and the error is smaller. To balance the fitting degree and error of short-term traffic flow prediction model, this neural network model LSTM combined with SVR is optimized by GA, and the short-term traffic flow prediction model based on LSTM-GASVR is obtained.

The structure of LSTM-GASVR is shown in Figure 16. The neural network is partly composed of the input layer, two layers of LSTM, dropout layer, and full connection layer. The output of the neural network is processed by the SVR to obtain the output results. Next moment workday flags  $h^m$  input to the first layer contain 64 LSTM networks and get 64 features. The 64 features and  $v_0$ , which is the traffic volume of period 0 to form a new vector, then input to the second layer in the network and get 64 features. These characteristics and traffic flow data of the next time step constitute a new vector input to 64 LSTM networks and get 64 LSTM characteristics.

And so on, the first LSTM layer outputs  $(m + 1) * 64$  data as input of the second LSTM layer. The output from last time step of the second LSTM will input to the dropout layer; finally, the results will be output by the dense layer of the model. The average of the results and the traffic flow at the previous moment inputs to the SVR model and then outputs the final results. "A in Figure 16" is shown in Figure 17, it is LSTM unit structure.  $C_{t-1}$  is the cell state,  $h_{t-1}$  is the output of the last LSTM, and  $X_t$  is the input of LSTM.  $f_b$ ,  $i_b$ , and  $o_t$  are output of forget gate, input gate, and output gate.  $\sigma$  and  $\tanh$  is activation function.  $C_t$  is the new cell state, and  $h_t$  is the output of LSTM.  $t-1$  means the last moment, and  $t$  is the current moment.

The LSTM-GASVR prediction process is shown in Figure 18. The input traffic data are preprocessed using MinMaxScaler normalization. According to the current short-term traffic flow data affected by the previous  $m$  data

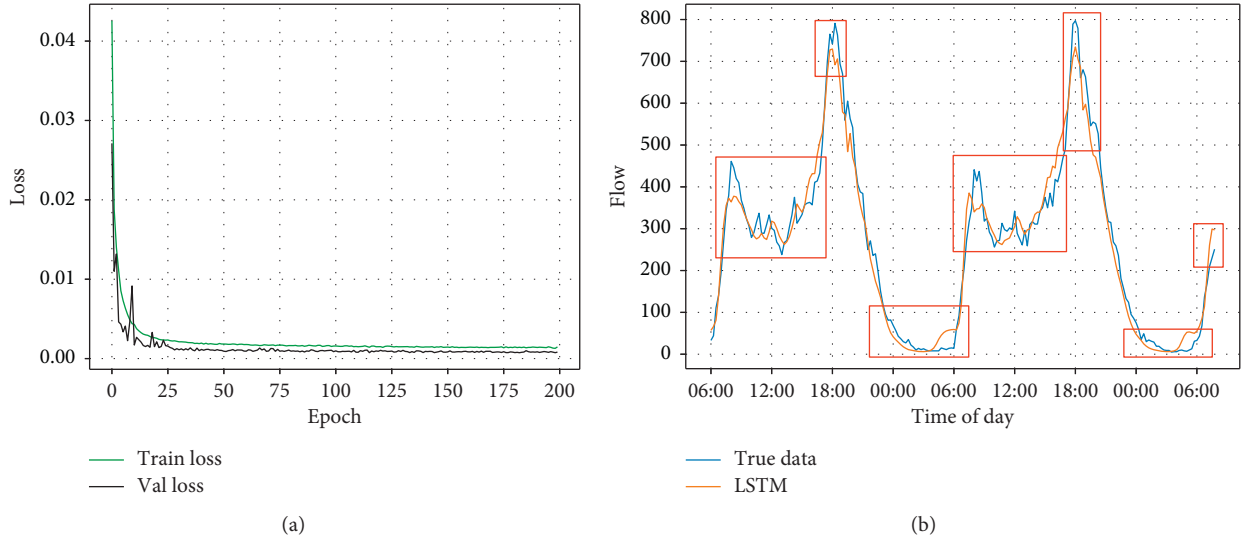


FIGURE 13: 24 steps LSTM training and prediction. (a) 24 steps LSTM loss function and (b) 24 steps LSTM prediction results.

TABLE 5: Comparison of short-term traffic flow prediction effect in different time steps.

| Time step | $R^2$ | Explanatory variance | MAPE  | MAE    | RMSE   |
|-----------|-------|----------------------|-------|--------|--------|
| 4         | 0.976 | 0.976                | 0.285 | 24.134 | 32.012 |
| 8         | 0.976 | 0.979                | 0.177 | 22.846 | 31.753 |
| 12        | 0.979 | 0.980                | 0.155 | 22.304 | 30.140 |
| 16        | 0.983 | 0.983                | 0.153 | 19.758 | 28.595 |
| 20        | 0.965 | 0.966                | 0.329 | 28.814 | 38.456 |
| 24        | 0.963 | 0.963                | 0.394 | 30.845 | 39.793 |

TABLE 6: Comparison of short-term traffic flow prediction effect of different prediction algorithms.

| Algorithm  | $R^2$ | Explanatory variance | MAPE  | MAE    | RMSE   |
|------------|-------|----------------------|-------|--------|--------|
| LSTM       | 0.983 | 0.983                | 0.153 | 19.758 | 28.595 |
| GRU        | 0.982 | 0.982                | 0.206 | 20.781 | 27.233 |
| CNN        | 0.952 | 0.962                | 0.293 | 33.297 | 45.206 |
| SAE        | 0.976 | 0.976                | 0.374 | 24.813 | 31.744 |
| ARIMA      | 0.975 | 0.975                | 0.222 | 23.313 | 32.381 |
| GASVR      | 0.966 | 0.967                | 0.144 | 26.092 | 37.995 |
| LSTM-GASVR | 0.982 | 0.982                | 0.118 | 20.156 | 28.237 |

and the workday flag, the array is reconstructed to  $[m + 1 : 1]$ . The data are divided into training set and test set, training set is used to train the model, and test set is used to test the prediction effect of the model. The data input to the neural network composed of LSTM, dropout, and dense, and then training of neural network. And the test data input neural network to make the preliminary prediction. The prediction results are renormalized, and the data are reset to the average of the LSTM model prediction results and the traffic flow at the previous time. The average input SVR model then optimizes the parameters of the model by the GA. After that, SVR model is trained according to the optimal parameters.

Finally, input the test data into the LSTM-GASVR model for prediction, and output the prediction results.

The prediction effect of the LSTM-GASVR model is shown in Figure 19. The time interval of the model is 15 minutes and the time step is 16. The green line in the figure represents the real traffic flow, and the red line represents the LSTM-GASVR model prediction value. It can be seen the prediction results of the LSTM-GASVR model are very close to the real value, which shows slightly higher error of prediction results at peak, but the error is small on the whole.  $R^2$  is 0.982, *explanatory variance* is 0.982, and MAPE is 0.118, and these are evaluation indicators of the model.

## 5. LSTM Prediction Model Parameter Determination

The batch size is 64, the epoch is 200, and the loss function is “rmsprop” as quantitative in the training of the LSTM model. The optimal performance of the model under different sampling time intervals and time steps is analyzed to determine the optimal sampling interval and time steps.

### 5.1. LSTM Prediction Model Sampling Interval Determination.

LSTM network model predicts the short-term traffic flow data with different time interval downsampling; then, we compare and analyze the training speed, loss function value, and the performance of the prediction results under different sampling intervals to determine the most suitable LSTM network model under the sampling interval.

The LSTM, trained with 5 min downsampling data, is shown in Figure 2(a). The loss function reaches a stable value of 0.002 at 25 epochs. As shown in Figure 2(b), the part marked by the red boxes is the larger part of the prediction difference. From Figure 2, the 5 min LSTM model is trained quickly, but the error of prediction results is large.

The LSTM of the trained 10 min downsampling data is shown in Figure 3(a). The loss function reaches a stable value

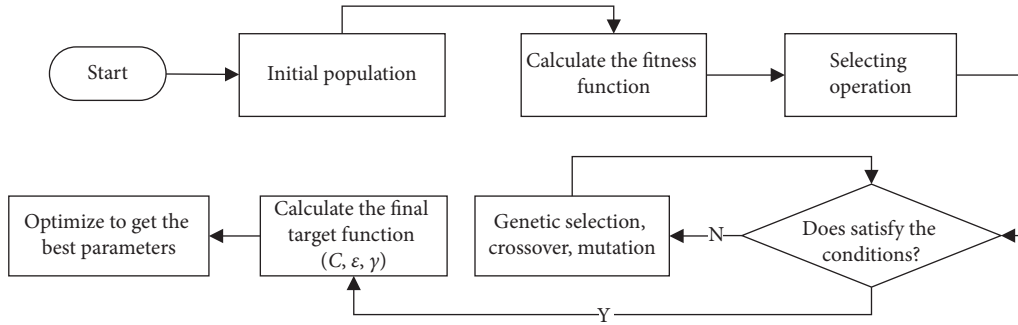


FIGURE 14: Flowchart of genetic algorithm.

TABLE 7: Probability interval.

| Individual                 | 1        | 2           | 3          | 4        |
|----------------------------|----------|-------------|------------|----------|
| Probability of individuals | 0.2      | 0.25        | 0.35       | 0.2      |
| Probability interval       | [0, 0.2) | [0.2, 0.45) | [0.45, 08) | [0.8, 1) |

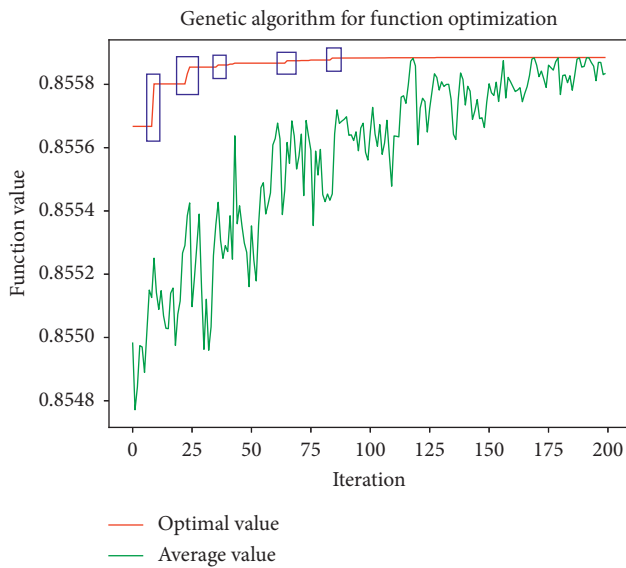


FIGURE 15: GA training.

of 0.002 at 50 epochs. As shown in Figure 3(b), the prediction difference is mostly labeled in the red boxes. From Figure 3, we can see that the LSTM model training speed of 10 min data is slow, and there are some deviations from the prediction results.

The LSTM of the analyzed 15 min downsampling data training is shown in Figure 4(a), and the loss function reaches a stable value of 0.002 at 50 epochs. As shown in Figure 4(b), the part marked in the red boxes is a larger part of the prediction difference. It can be seen from the figure that the 15 min LSTM model has a fast training speed, and the prediction results are generally good, and the errors at the peak and fluctuation are larger.

The LSTM of 20 min downsampling data training is analyzed. As shown in Figure 5(a), the epoch is 75 where the loss function is stable at 0.003. As shown in Figure 5(b), the prediction difference is mostly marked in red boxes. From

Figure 5, the 20 min LSTM model training speed is better, the value of loss function is slightly higher, the prediction results are general, and the prediction error is large at peak and fluctuation.

The LSTM model obtained from the training of 25 min downsampling data is shown in Figure 6(a). The stable value of the loss function is 0.003 at 75 epochs. As shown in Figure 6(b), the part marked by the red boxes is the larger part of the prediction difference. From Figure 6 the 25 min LSTM model training speed is slightly slow, the value of the loss function is slightly higher, the prediction results are general, and the prediction error is large at the peak and the fluctuation.

The LSTM, trained under 30 min downsampling data, is shown in Figure 7(a). The loss function reaches a stable value of 0.003 at 75 epochs. As shown in Figure 7(b), the red boxes indicate that the prediction results showed slightly larger error. From the figure, the 30 min LSTM model training speed of the data is slightly slower, the loss function value is higher, the prediction results are general, and the prediction result error is larger at peak and fluctuation.

Since the difference judgment of the fitting curve is difficult, the prediction results of the LSTM network model with different sampling intervals are again compared and analyzed by various evaluation indexes. As shown in Table 4, the model evaluation trained by the data of 15 min is the best.

### 5.2. LSTM Prediction Model Time Step Determination.

The LSTM model predicts the short-term traffic flow data of 15 min downsampling with different time steps  $n$  (the current moment is affected by the previous  $n$  moment). The most suitable time step is selected by analyzing the stability of the training process, the value of the loss function, and the curve of the fitting degree of the predicted results to the real value.

The LSTM model is trained with 15-minute interval downsampling data at 4 time steps. Its loss function image is

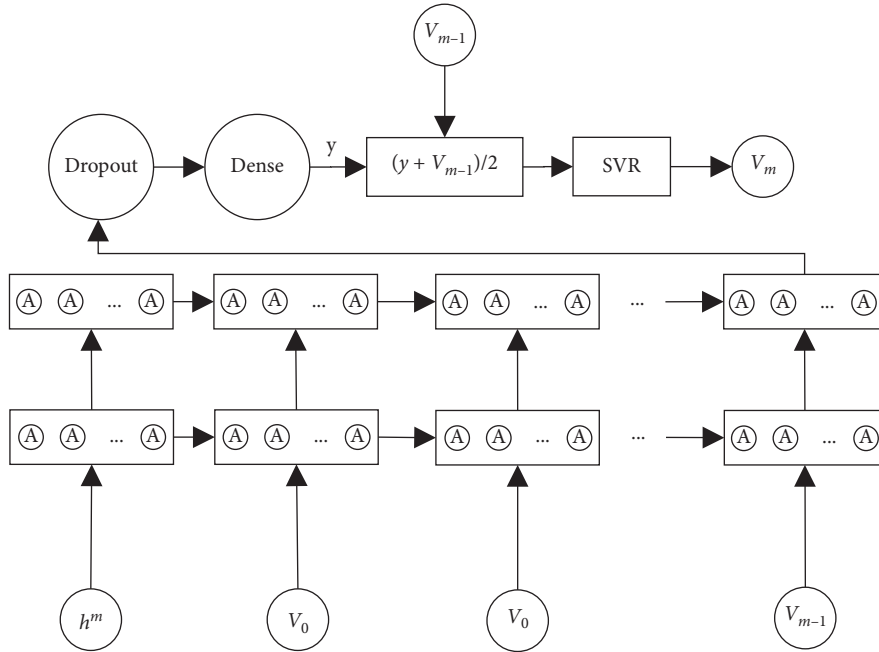


FIGURE 16: LSTM-GASVR structure.

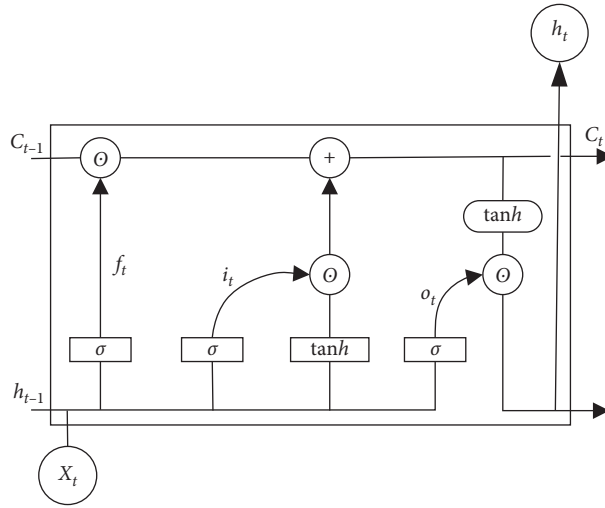


FIGURE 17: The structure of LSTM.

shown in Figure 8(a), and the loss function is stable around 0.002 at 50 epochs. The prediction results are shown in Figure 8(b). The red boxes are marked with a large prediction difference. The training speed of the LSTM model with 4 time steps is slightly slower, the loss function value is higher, the prediction results are general, and the prediction result error at peak is very high.

Analyzing the 15 min of downsampling data trained at 8 time steps LSTM, as shown in Figure 9(a), the loss function is stabilized around 0.003 at 75 epochs. As shown in Figure 9(b), the part marked in the red boxes shows that the prediction difference is large. The training speed of the LSTM model with 8 time steps is slow, the loss function value is slightly larger, the overall prediction results are

general, and the prediction results at peak and fluctuation are obviously poor.

The LSTM model of downsampling data trained at 12 time steps and the loss function are shown in Figure 10(a), the loss function reaches a stable value of 0.002 at 50 epochs, the prediction results are shown in Figure 10(b), and the prediction difference is mostly marked in red boxes. It can be seen from the figure that the LSTM model with 12 time steps has a fast training speed, accurate prediction results, but a large error in the prediction results at the peak and with large fluctuations.

The LSTM model, trained with 15 min interval downsampling data at 16 time steps, is shown in Figure 11(a), and the loss function reaches a stable value of 0.003 at 50 epochs. As shown in Figure 11(b), the red boxes are marked with a

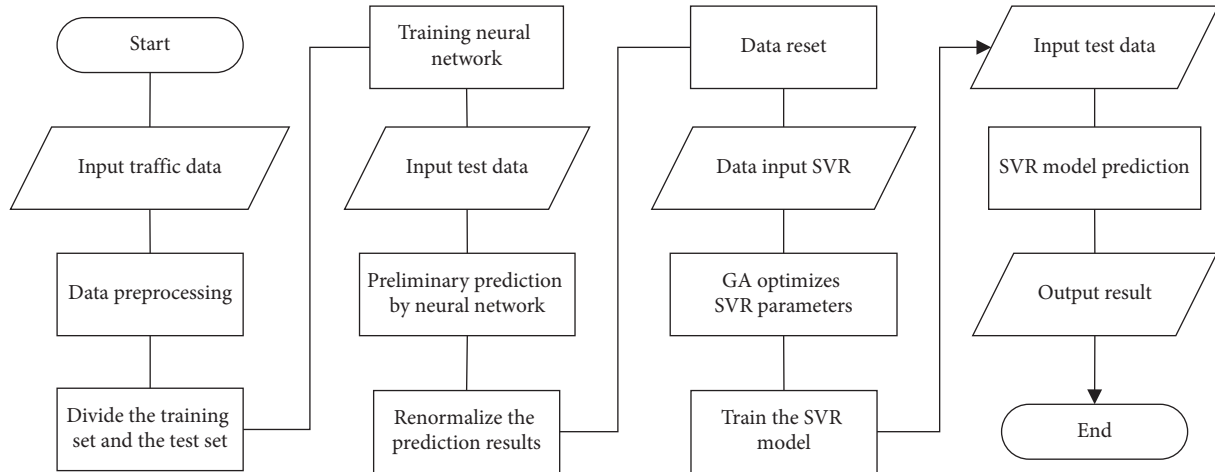


FIGURE 18: LSTM-GA SVR prediction process.

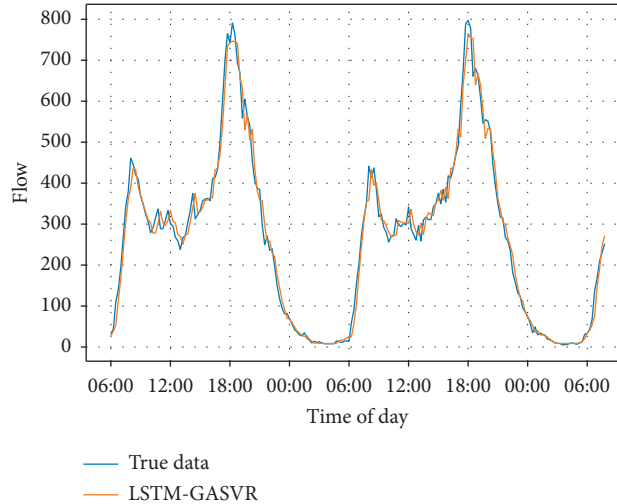


FIGURE 19: LSTM-GASVR prediction results.

large difference in prediction. The LSTM model with 16 time steps shows better training speed, and the loss function value is low. Overall, the prediction results are accurate, but slightly higher error of prediction results at peak.

The LSTM with downsampling interval of 15 min was trained at 20 time steps. As shown in Figure 12(a), the loss function at 50 epochs is stable at around 0.002. As shown in Figure 12(b), the prediction difference of red boxes part is large. The training speed of LSTM model with 20 time steps is general, the prediction results are general, and the error of peak, fluctuation, and valley bottom part is very high.

The LSTM model were trained of the data sampled at intervals of 15 min at 24 time steps, and the loss function is shown in Figure 13(a), reaching a stable value of 0.003 at 50 epochs. As shown in Figure 13(b), the part marked by the red boxes is the larger part of the prediction difference. From the figure, the training speed of LSTM model with 24 time steps is general, and the predicted results are general. The prediction error is large at the peak and the fluctuation, and the prediction results of the valley bottom are very high too.

The LSTM model predicts the short-term traffic flow data of 15 min downsampling with different time steps  $n$  (the current moment is affected by the previous  $n$  moment). Then, we compare the prediction results according to various evaluation indexes. As shown in Table 5, it is found that the prediction effect of 16 time steps is the best.

The optimal performance of the LSTM model under different sampling time intervals and time steps is analyzed, and the sampling interval is finally determined to be 15 min at 16 time steps. The LSTM model training speed is the fastest at this time, the loss function reaches the stable value at 25 epochs and the loss function value is small, and the prediction accuracy is the highest.

## 6. Comparative Analysis of Short-Time Traffic Flow Prediction Model

Through LSTM, GRU (gated recurrent unit), CNN (convolutional neural networks), SAE (stacked autoencoder), ARIMA (auto regressive integrated moving average), SVR,

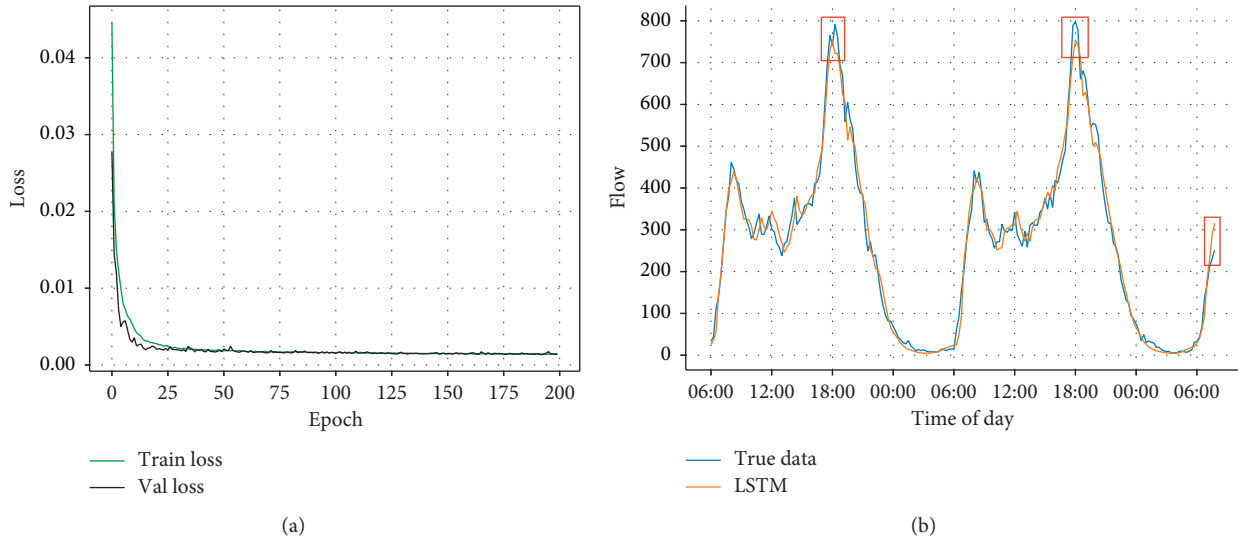


FIGURE 20: LSTM training and prediction. (a) LSTM loss function and (b) LSTM prediction results.

LSTM-GASVR prediction of 15 minutes short time traffic volume, in this paper, the training speed and loss function of LSTM, GRU, CNN, and SAE in training are compared and analyzed, the prediction results of seven algorithms are compared to fit the curve of real value, and the prediction effect is analyzed.

Analyzing the LSTM model trained with the downsampling interval of 15 min data at the 16 time steps, as shown in Figure 20(a), the loss function reaches a stable value of 0.002 at 25 epochs. As shown in Figure 20(b), the part marked in the red boxes is the larger part of the prediction difference, the training speed of the LSTM model is faster, the prediction results are generally more accurate, and the error of the prediction results at the peak is slightly larger.

The data sampled by 15 min are trained to GRU model at the 16 time steps, and the loss function is stable around 0.002 at 75 epochs as shown in Figure 21(a). As shown in Figure 21(b), the partial prediction difference is marked by the red boxes. From the figure, we can see that the model training speed is general, the prediction accuracy is general, and the prediction result error at the peak is very large.

The 1D CNN model of 15 min downsampling data trained at 16 time steps is analyzed. As shown in Figure 22(a), the 50-epoch training sets get a stable training set loss function value of 0.007, and the validation set loss function stability value of 0.003. As shown in Figure 22(b), the red boxes are marked with a large difference in prediction, the model training speed is general, the loss function value is large, the prediction accuracy is not high, and the prediction error is large.

The SAE model trained with 15 min downsampling data at 16 time steps is analyzed. The loss function reaches a stable value of 0.001 at 100 epochs, as shown in Figure 23(a). As shown in Figure 23(b), the difference in prediction is mostly marked in red boxes, the SAE model training speed is

average, the loss function value is small, the prediction accuracy is not high, and the prediction effect is very poor at peak, fluctuation, and turning point.

The ARIMA model and the GASVR model were used to predict 15 min of short-term traffic flow, respectively. The prediction results of the ARIMA model are shown in Figure 24, the red boxes are marked with a large part of the prediction difference, it can be seen that the prediction accuracy of peak value is low, and it takes a long time in the prediction process of the ARIMA model, which does not meet the real-time requirement of short-term traffic flow prediction. As shown in Figure 25, for the prediction results of the GASVR model, the red boxes are marked with a large part of the prediction difference. It can be seen that the prediction accuracy is slightly lower at the peak value, and the GASVR model prediction curve is shifted backward compared with the real value.

In Figure 26, the prediction results of the LSTM-GASVR model show that the model effectively improves the migration phenomenon of the GASVR and improves the accuracy of the LSTM model. These are some differences were marked by the red boxes, the prediction accuracy of the LSTM-GASVR model is not enough at the peak of traffic flow. But it has a little influence on the results of the algorithm, and the LSTM-GASVR model predicted the most accurate results.

The timeliness of the LSTM-GASVR model is normal compared with other algorithms. The prediction time of the model is 0.003 s longer than GASVR, and the time is 2 s shorter than ARIMA. Compared with the LSTM model, the prediction time of the LSTM-GASVR is 0.001 s longer than that of the LSTM model. In addition, compared with CNN, SAE, and GRU models, the LSTM-GASVR model takes time to predict that traffic flow is same almost.

However, the LSTM-GASVR model's timeliness is normal compared with other algorithms and the accuracy is



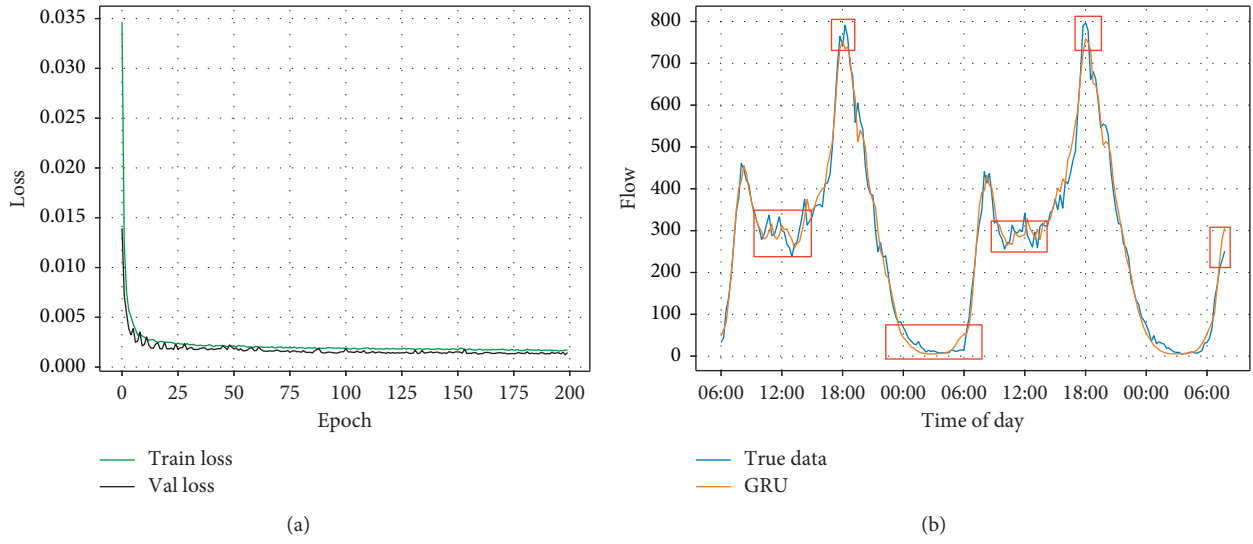


FIGURE 21: GRU training and prediction. (a) GRU loss function and (b) GRU prediction results.

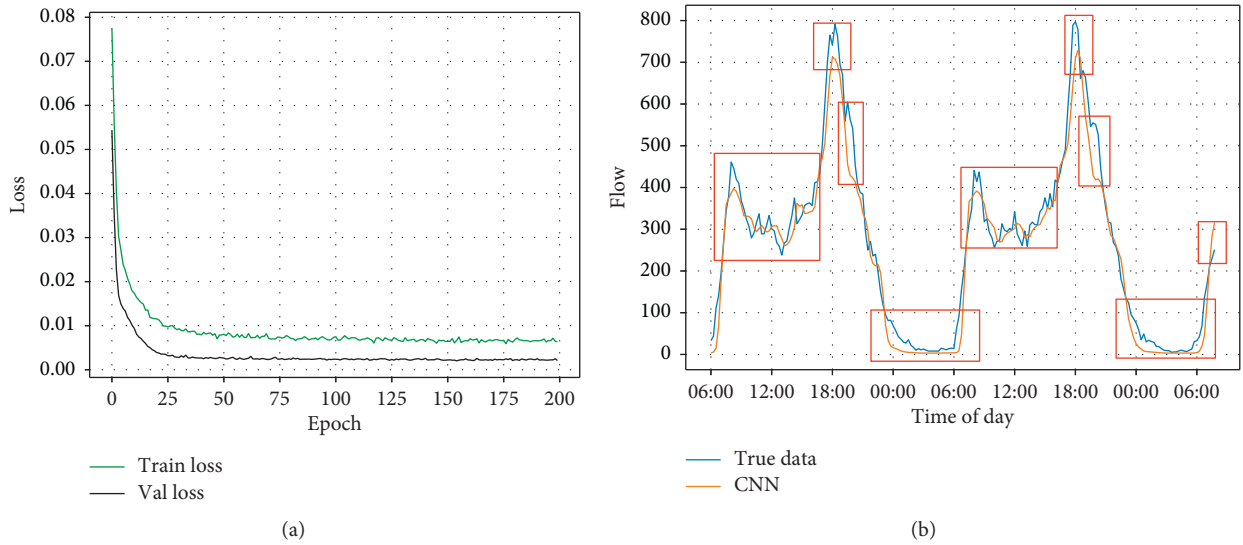


FIGURE 22: CNN training and prediction. (a) CNN loss function and (b) CNN prediction results.

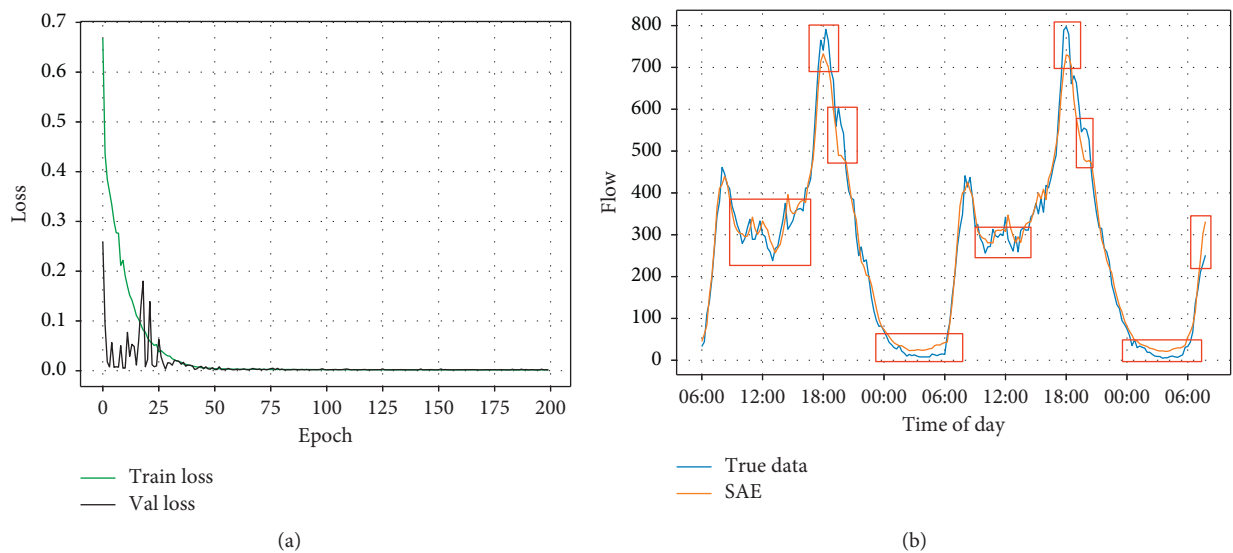


FIGURE 23: SAE training and prediction. (a) SAE loss function and (b) SAE prediction results.

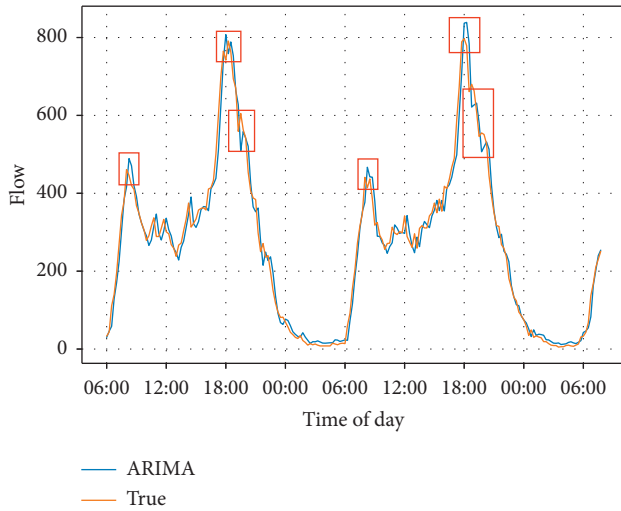


FIGURE 24: ARIMA prediction results.

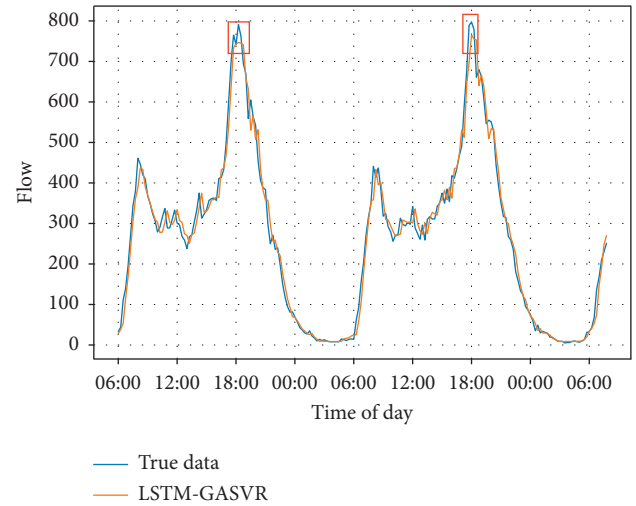


FIGURE 26: LSTM-GASVR prediction results.

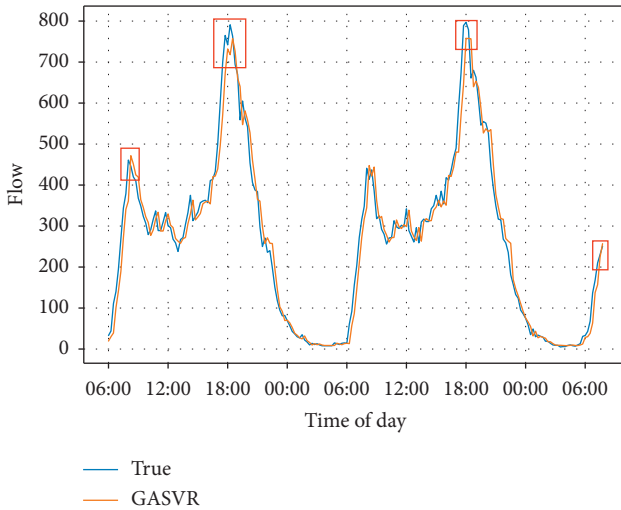


FIGURE 25: GASVR prediction results.

well. According to the above algorithm, the 15-minute downsampling data are predicted, and the prediction results are compared and analyzed according to various evaluation indexes. As shown in Table 6, it is found that the comprehensive prediction effect of the LSTM-GASVR model is the best.

Six prediction algorithms LSTM, GRU, CNN, SAE, GASVR, and LSTM-GASVR are analyzed and compared and the conclusions are summarized in Table 8.

## 7. Conclusion

Based on the charging data from May 2018 to May 2019 at a toll station around Shaanxi Province, the data normalization and reconstruction are adopted. And the working day flag bit is added to enhance the data dimension and to realize the data preprocessing of short-

TABLE 8: Analysis of short-time traffic flow prediction algorithm for different prediction algorithms.

| Algorithm  | Summary  |
|------------|--|
| LSTM       | The prediction effect is good and timeliness is general, but there are big differences for different models trained  |
| GRU        | Simpler than the LSTM structure, but the traffic flow prediction at the turning point is less effective. It is suitable for predicting stable traffic flow, and the prediction effect is not good when the volatility is large |
| CNN        | SAE model has large prediction error when traffic flow data is small   |
| SAE        | ARIMA model prediction error is general and timeliness is poor   |
| ARIMA      | The prediction effect is good and timeliness is better, but there is a certain shift phenomenon  |
| GASVR      | The timeliness of the model is normal, and a variety of time interval data prediction effect is better and more stable   |
| LSTM-GASVR |  |

term traffic flow. Comparing and analyzing the evaluation indexes of various time series, an effective combination evaluation index of short-term traffic flow prediction is established. Through analyzing the neural network prediction model and other machine learning prediction models and using GA to optimize the SVR model parameters, a short-term traffic flow prediction model based on LSTM-GASVR is proposed. By analyzing and comparing different time intervals in the multiple groups of experimental results, we selected 15 min as the time interval, and the time step is 16. This model is used to predict the short-term traffic flow data, and various prediction models are analyzed by combination index. The LSTM-GASVR model has normal timeliness and the best and stable prediction effect,  $R^2$  is 0.982, *explanatory variance* is 0.982, and *MAPE* is 0.118. The next step is to optimize the prediction accuracy at the peak traffic volume.

## Data Availability

The traffic flow data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the National Key Research and Development Program of China under Grant 2018YFB1600600, the National Natural Science Funds of China under Grant 51278058, “111 Project on Information of Vehicle-Infrastructure Sensing and ITS” under Grant B14043, Shaanxi Natural Science Basic Research Program under Grants 2019NY-163 and 2020GY-018, Joint Laboratory for Internet of Vehicles, Ministry of Education-China Mobile Communications Corporation under Grant 213024170015, the Special Fund for Basic Scientific Research of Central Colleges, Chang’an University in China under Grants 300102329101 and 300102249101.

## References

- [1] C. Zheng, *A Study on Prediction Model of Urban Road Traffic Flow in Short Time*, University of Science and Technology of China, Hebei, China, 2016, in Chinese.
- [2] B. Lu, S. Qin, and G. Ma, “Forecast of short-term traffic flow based on multi-source traffic data fusion,” *Journal of Chongqing Jiaotong University (Natural Science Edition)*, vol. 38, no. 5, pp. 13–19, 2019, in Chinese.
- [3] A. Miglani and N. Kumar, “Deep learning models for traffic flow prediction in autonomous vehicles: a review, solutions, and challenges,” *Vehicular Communications*, vol. 20, 2019.
- [4] F. Van Wageningen-Kessels, H. van Lint, K. Vuik, and S. Hoogendoorn, “Genealogy of traffic flow models,” *EURO Journal on Transportation and Logistics*, vol. 4, no. 4, pp. 445–473, 2015.
- [5] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, “Short-term traffic forecasting: where we are and where we’re going,” *Transportation Research Part C Emerging Technologies*, vol. 43, no. 1, pp. 3–19, 2014.
- [6] A. Ermagun and D. Levinson, “Spatiotemporal traffic forecasting: review and proposed directions,” *Transport Reviews*, vol. 38, no. 6, pp. 786–814, 2018.
- [7] W. Luo, B. Dong, and Z. Wang, “Short-term traffic flow prediction based on CNN-SVR hybrid deep learning model,” *Transport Systems Engineering and Information*, vol. 17, no. 5, pp. 68–74, 2017, in Chinese.
- [8] J. Tang, X. Chen, Z. Hu, F. Zong, C. Han, and L. Li, “Traffic flow prediction based on combination of support vector machine and data denoising schemes,” *Physica A: Statistical Mechanics and its Applications*, vol. 534, Article ID 120642, 2019.
- [9] L. N. N. Do, H. L. Vu, B. Q. Vo, Z. Liu, and D. Phung, “An effective spatial-temporal attention based neural network for traffic flow prediction,” *Transportation Research Part C*, vol. 108, pp. 12–28, 2019.
- [10] H. Wang, L. Liu, S. Dong, Z. Qian, and H. Wei, “A novel work zone short-term vehicle-type specific traffic speed prediction model through the hybrid EMD-ARIMA framework,” *Transportmetrica B: Transport Dynamics*, vol. 4, no. 3, pp. 159–186, 2016.
- [11] S. Vasantha Kumar and L. Vanajakshi, “Short-term traffic flow prediction using seasonalARIMA model with limited input data,” *European Transport Research Review*, vol. 7, no. 3, 2015.
- [12] Y. Rajabzadeh, A. H. Rezaie, and H. Amindavar, “Short-term traffic flow prediction using time-varying Vasicek model,” *Transportation Research Part C*, vol. 74, pp. 168–181, 2017.
- [13] A. Salamanis, G. Margaritis, D. D. Kehagias, and G. Matzoulas, D. Tzovaras, Identifying patterns under both normal and abnormal traffic conditions for short-term traffic prediction,” *Transportation Research Procedia*, vol. 22, pp. 665–674, 2017.
- [14] R. Neuhold, F. Garolla, S. Oliver, and M. Fellendorf, “Predicting and optimizing traffic flow at toll plazas,” *Transportation Research Procedia*, vol. 37, pp. 330–337, 2019.
- [15] R. Liu, F. Hong, C. Lu, and W. Jiang, “Short-term traffic flow prediction based on deep circulation neural network,” *Journal of Physics: Conference Series*, vol. 1176, no. 3, Article ID 032020, 2019.
- [16] L. Qu, W. Li, W. Li, D. Ma, and Y. Wang, “Daily long-term traffic flow forecasting based on a deep neural network,” *Expert Systems With Applications*, vol. 121, pp. 304–312, 2019.
- [17] L. Zhi, *Study on the Prediction of Expressway Traffic Flow Based on BP Neural Network*, Wuhan University of Technology, Wuhan, China, 2014, in Chinese.
- [18] Li Chin-ming, *A Study on the Prediction Technology of Traffic Flow at Crossroads Facing Intelligent Transportation*, Beijing University of Posts and Telecommunications, Beijing, China, 2017, in Chinese.
- [19] J. Kang, Z. Duan, L. Tang et al., “A short-time traffic flow prediction method for smoothing,” *Measurement And Control Technology*, vol. 37, no. 2, pp. 33–37, 2018, in Chinese.
- [20] S. Wang, Y. Gu, Li Meng et al., “Short-term traffic flow prediction of expressway based on chaos theory and MEA-BPNN model,” *Shandong Science*, vol. 32, no. 2, pp. 98–107, 2019, in Chinese.
- [21] J. Xue, *Research and Implementation of Traffic Information Prediction Model Based on Deep Learning*, University of Information Engineering, Strategic Support Force, Henan, China, 2018, in Chinese.
- [22] Z. Leng, J. Gao, Y. Qin, X. Liu, and J. Yin, “Short-term forecasting model of traffic flow based on GRNN,” in *Proceedings of the 25th Chinese Control and Decision Conference (CCD)*, pp. 3816–3820, IEEE, Guiyang, China, 2013.
- [23] Y. Jia, J. Wu, and Y. Du, “Traffic speed prediction using deep learning method,” in *Proceedings of the IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1217–1222, IEEE, Rio de Janeiro, Brazil, 2016.
- [24] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, “Long short-term memory neural network for traffic speed prediction using remote microwave sensor data,” *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 187–197, 2015.
- [25] Y. Jia, J. Wu, M. Ben-Akiva, R. Seshadri, and Y. Du, “Rainfall-integrated traffic speed prediction using deep learning method,” *IET Intelligent Transport Systems*, vol. 11, no. 9, pp. 531–536, 2017.
- [26] Y. Wu and H. Tan, “Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework,” 2016, <https://arxiv.org/abs/1612.01022>.
- [27] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, “LSTM network: a deep learning approach for short-term traffic forecast,” *IET Intelligent Transport Systems*, vol. 11, no. 2, pp. 68–75, 2017.
- [28] W. Wangyang, H. Wu, and H. Ma, “An auto encoder and LSTM-based traffic flow prediction method,” *Sensors*, vol. 19, no. 13, p. 2946, 2019.

## Research Article

# Research on Complex Classification Algorithm of Breast Cancer Chip Based on SVM-RFE Gene Feature Screening

Guobin Chen,<sup>1</sup> Xianzhong Xie,<sup>1</sup> and Shijin Li<sup>1,2</sup> 

<sup>1</sup>College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, China

<sup>2</sup>Academic Affairs Office, Yunnan University of Finance and Economics, Kunming 650221, Yunnan, China

Correspondence should be addressed to Shijin Li; [shijin\\_lee@ynufe.edu.cn](mailto:shijin_lee@ynufe.edu.cn)

Received 1 April 2020; Revised 10 May 2020; Accepted 23 May 2020; Published 13 June 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Guobin Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Screening and classification of characteristic genes is a complex classification problem, and the characteristic sequences of gene expression show high-dimensional characteristics. How to select an effective gene screening algorithm is the main problem to be solved by analyzing gene chips. The combination of KNN, SVM, and SVM-RFE is selected to screen complex classification problems, and a new method to solve complex classification problems is provided. In the process of gene chip pretreatment, LogFC and  $P$  value equivalents in the gene expression matrix are screened, and different gene features are screened, and then SVM-RFE algorithm is used to sort and screen genes. Firstly, the characteristics of gene chips are analyzed and the number between probes and genes is counted. Clustering analysis among each sample and PCA classification analysis of different samples are carried out. Secondly, the basic algorithms of SVM and KNN are tested, and the important indexes such as error rate and accuracy rate of the algorithms are tested to obtain the optimal parameters. Finally, the performance indexes of accuracy, precision, recall, and F1 of several complex classification algorithms are compared through the complex classification of SVM, KNN, KNN-PCA, SVM-PCA, SVM-RFE-SVM, and SVM-RFE-KNN at  $P = 0.01, 0.05, 0.001$ . SVM-RFE-SVM has the best classification effect and can be used as a gene chip classification algorithm to analyze the characteristics of genes.

## 1. Introduction

Since the birth of gene chip technology, a large number of feature selection methods for gene expression microarray data have emerged in academia. Most of these methods focus on the quality of the selected genes, while few people pay attention to the efficiency of the algorithm itself. Gene expression microarray data has a large number of characteristic genes. If it is not an efficient characteristic selection method, the whole process of key gene selection will become very long. Many existing classical feature selection methods have low efficiency, and some even reach unacceptable levels. Among them, representatives include CFS, mRMR, and SVM-RFE. Especially in SVM-RFE, the whole selection process is very time-consuming. Some researchers have also improved SVM-RFE, but the time-consuming problem has not been fundamentally improved. This chapter takes SVM-RFE as the research object, and SVM and RFE are improved, respectively. By

introducing a more efficient implementation of the classical linear support vector machine to reduce the time consumption of the basic feature selection process and proposing a recursive feature elimination strategy with variable step size to reduce the iteration times of the basic feature selection process, the combination of the two finally attempts to fundamentally solve the inefficiency problem of SVM-RFE.

A support vector machine based on recursive feature eligibility (SVM-RFE) [1] was proposed by Guyon et al. In 2002. This method makes full use of the characteristics of SVM; that is, it can rank and score all genes according to their importance while training the SVM classification method and combine the recursive feature elimination strategy to make feature selection. Duan et al. [2] improved SVM-RFE to deal with only two classification problems and proposed a one-to-one and one-to-many multiclassification SVM-RFE method, which enables SVM-RFE to deal with multiclassification problems.

Aiming at the low efficiency of SVM-RFE feature selection process, Ding and Wilkins [3] improved the iterative process of RFE, from deleting one feature at a time to deleting several, which improved the efficiency of the algorithm without losing the classification accuracy. Yoon and Kim [4] proposed a SVM-RFE method based on mutual information, which solves the problem that the SVM-RFE method does not consider feature correlation in the process of feature selection to a certain extent. Tang et al. [5] divided SVM-RFE into two stages. In the first stage, rough selection is carried out on the features to be selected, filtering out irrelevant features, redundant features, and noise features. In the second stage, finer feature selection is carried out on the basis of the first stage. The next two sections will introduce SVM-RFE in detail and analyze the reasons for its inefficiency in depth. Tang et al. [6] feature clustering SVM-RFE (FCSVM-RFE) feature clustering to enhance SVM-RFE gene selection. The proposed method first roughly selects genes and then ranks the selected genes. Clustering algorithm is used to cluster genes into genomes, in which each gene has a similar expression profile. SVM-RFE was used to rank these representative genes. FCSVM-RFE reduces computational complexity and redundancy. Although SVM-RFE can effectively delete irrelevant functions, it cannot handle most redundant functions [7]. In order to overcome this shortcoming, this paper develops a new feature selection method, the core of which is to delete redundant features according to the correlation between features before using SVM-RFE. The proposed method was tested on a pancreatic cancer microarray dataset. The method is much better than baseline SVM-RFE in classification accuracy. In order to improve the accuracy of classification, radial basis function (RBF) kernel is also introduced [8]. Chen and Zhu [9] proposed a feature selection method based on support vector machine recursive feature elimination (SVM-RFE) and binary particle swarm optimization (BPSO) algorithm. SVM-RFE removes some irrelevant features to reduce the data dimension and then continues to search for the best subset and uses some better SVM-RFE subsets as part of the initial PSO population and has a good starting point. SVM-RFE not only reduces the search space of particles but also provides prior experience, thus improving the search efficiency and accuracy of the algorithm. Anaissi et al. [10] used ESVM recursive feature elimination (ESVM-RFE) for gene selection. It follows the concepts of integration and bagging used in random forest but adopts backward elimination strategy, which is the basic principle of RFE algorithm. The principle behind this is that using randomly drawn boot program samples from the training set to build an integrated SVM model will generate different feature levels, which will then be summarized into one feature level. As a result, the decision to delete features is based on the ranking of multiple SVM models, rather than selecting a specific model. However, in the classification of unbalanced datasets, imbalance is a common problem in gene expression microarray data [11]. Generally speaking, people are only interested in a few categories because the few categories are usually patients, while normal people often account for the majority. For the classification method, too few samples in a certain category means that the category

contains less information, so the classification model finally learned by the classification algorithm can easily predict patients among normal people when making classification prediction [12]. Especially for small sample data such as gene expression microarray data, it becomes more important to solve the problem of category imbalance. The most basic methods to solve the problem of category imbalance are upsampling and downsampling. Zhou and Wang [13] proposed a feature selection method combining relief-F and SVM-RFE algorithm. The algorithm integrates the weight vector from relief-F into the SVM-RFE method. In this method, relief-F filters out many noisy functions in the first stage. Then, a new sorting criterion based on the SVM-RFE method is applied to the final feature subset. A SVM classifier is used to evaluate the final image classification accuracy. A new method for multiclass gene selection and classification based on multiple supports vector machines recursive feature elimination (SVM-RFE) is proposed [14]. For the multiclass DNA microarray problem, we solve it as a multibinary classification problem. The "one-to-all" method is used to decompose multiple types of tasks into multiple binary problems, and SVM-RFE selects genes for each binary problem. The SVM classifier is used to train selected gene data for binary problems. Firstly, the basic method of SVM is introduced, and the application of RFE algorithm is explained in detail. Secondly, the chip GSE76275 screens and classifies different  $P$  values under SVM-RFE algorithm. Finally, the classification effect of SVM-RFE algorithm after filtering with different  $P$  values is illustrated by comparative research under different SVM-RFE-KNN, SVM-RFE-SVM, and other four algorithms.

## 2. Relevant Theoretical Works

*2.1. Support Vector Machine.* A support vector machine (SVM) is recognized as one of the most classical machine learning algorithms. Its essence is the maximum interval classification method. At this time, a support vector machine can only deal with linearly separable data classification problems and is called hard interval support vector machine. Soft interval support vector machine was proposed in 1995. At this time, a support vector machine can deal with data classification problems that are approximately linearly separable. Subsequently, support vector machines have been further developed. Support vector machines, support vector regression machines, and multiclassification support vector machines based on kernel techniques have been proposed one after another. At this time, the support vector machine has formed a very complex and complete theoretical system, which can not only deal with linear separable problems but also classify nonlinear separable data, becoming very powerful. The SVM-RFE algorithm uses a support vector machine based on linear kernel. The support vector machine model is shown in Figure 1.

The algorithm idea of the support vector machine is actually very simple. For hard interval support vector machines, the whole process is divided into three steps: first, the dataset is linearly separable; second, finding two hyperplanes requires that no data points fall between the two planes.

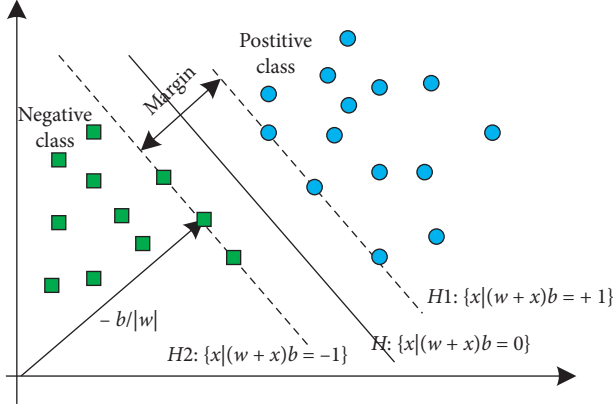


FIGURE 1: Support vector machine model.

Third, maximize the distance between the two planes. The objective function at this time is

$$\max \frac{2\eta}{\|w\|} \quad (1)$$

$$\text{s.t. } y_i(w \cdot x_i + b) \geq \eta, \quad i = 1, \dots, n,$$

where  $(2\eta/\|w\|)$  is the distance between the two hyperplanes and is the target optimization value,  $x_i$  and  $y_i$  represent the  $i$ -th sample and the corresponding label, respectively, and  $y_i(w \cdot x_i + b)$  represents the distance from the point  $x_i$  to the nearest hyperplane.  $y_i(w \cdot x_i + b) \geq \eta$  means that point  $x_i$  cannot fall between two hyperplanes. In order to facilitate the solution, formula (1) is usually transformed into a quadratic programming problem:

$$\min \frac{1}{2} \|w\|^2 \quad (2)$$

$$\text{s.t. } y_i(w \cdot x_i + b) \geq 1, \quad i = 1, \dots, n.$$

In all classifications, the classification interval of the optimal plane is the largest; at this time,  $\|w\|^2$  is the smallest, H is called the optimal classification line, and the training samples on H1 and H2 are called support vectors. The Lagrange optimization method is used to obtain it. Assuming that  $b = (b_1, b_2, \dots, b_n)$  and equation (2) constitute Lagrange multiple terms, the maximum value is taken.

$$W(a) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{j=1}^n a_i a_j y_i y_j x_i x_j, \quad (3)$$

where  $a_i \geq 1$ ,  $\sum_{i=1}^n y_i a_i = 0$  is quadratic programming that can optimize equation (3). Assuming that there is a maximum vector  $a^0 = (a_1^0, a_2^0, \dots, a_n^0)$  of equation (3) and the optimal hyperplane is described by  $(w_0, b_0)$ , then  $w_0$  is shown in the following equation:

$$w_0 = \sum_{i=1}^n a_i^0 y_i x_i. \quad (4)$$

If the restriction condition is proposed in equation (4), the decision function of the optimal classification is shown in the following equation:

$$f(x) = \text{sgn} \left( \sum_{i=1}^n a_i^0 y_i x_i + b_0 \right) = 0. \quad (5)$$

Equation (5) introduces Lagrange's equation:

$$L(w, b, a) = \frac{1}{2} (w \cdot w) - \sum_{i=1}^n a_i y_i (w x_i + b) - 1, \quad (6)$$

where  $a$  is the Lagrange coefficient. By differentiating  $w$  and  $b$ , we obtain the quadratic programming problem:

$$\begin{cases} \min & \frac{1}{2} \sum_{j=1}^n a_i a_j y_i y_j [\phi(x_i) \cdot \phi(x_j)] - \sum_{i=1}^n a_i \\ \text{s.t.} & \sum_{i=1}^n y_i a_i = 0 \\ & a_i \geq 0, \quad (i = 1, 2, \dots, n). \end{cases} \quad (7)$$

**2.2. Recursive Feature Elimination (RFE).** The main idea of recursive feature elimination is to repeatedly build models (such as SVM or regression models). The importance of each feature is obtained through the attribute value returned by the learner or the importance score of the feature. Then, the least important feature variable is removed from the current feature set. Then, the model is constructed on the remaining characteristic variables. Repeat the abovementioned process until there is only one feature variable left. This process constructs a model of feature number minus one time. The order in which features are eliminated is the importance ranking of features. This is a greedy algorithm to find the optimal feature subset, which requires a lot of computation and requires high hardware requirements of computers. The stability of RFE depends to a large extent on which a model is selected at the bottom during iteration. For example, if the ordinary linear regression adopted by RFE is unstable without regularization, then RFE is unstable. If a ridge is used and the regression regularized by the ridge is stable, then RFE is stable. For example, a linear kernel support vector machine SVM-RFE, as an effective feature selection method, has been successfully applied to fault diagnosis. However, some problems may be nonlinear.

SVM-RFE is a supervised sequential backward selection algorithm. In the linear classifier, it takes the discriminant information of each feature to the objective function as the sorting coefficient. That is, the contribution of the weight vector to the classification surface  $y_i(w \cdot x_i + b)$  is used to construct the feature ranking table. If the weight corresponding to the feature is larger, the decision function will be affected more, and the feature with larger weight has more discrimination information. Each iterate removes a feature with the smallest weight and then retrains the classifier until the feature ranking table is completed. The sorting principle can also be analyzed from the objective function of the following formula:

Input: the training samples:  $X_0 = [x_1, x_2, \dots, x_i]^T$ ,  $x_i$  is the samples of d-dimensional space

Category label:  $y = [y_1, y_2, \dots, y_i]^T$

Initialization: feature sort  $r = []$ , current feature index sequence  $s = [1, 2, \dots, d]$

Feature sorting: iterates in a loop until  $s = []$ ;

Step 1: obtain a new data sample according to the current feature:  $X = X_0(:, s)$

Step 2: train SVM with a new sample set to obtain support vector related parameters:  $a = \text{SVMtrain}(X, y)$ ,  $X = X_0(:, s)$

Step 3: calculate the sorting factor

Step 4: find out the feature  $f = \arg \min(\text{Rank}(i))$  with the smallest sorting criterion and add it to the feature sorting table:

$r = [s(f), r]$

Step 5: remove the feature with the smallest sorting coefficient from the current remaining dataset:  $s = s(1: f - 1, f + 1: \text{length}(s))$

Output: sorted list of features.

ALGORITHM 1: SVM-RFE algorithm flow.

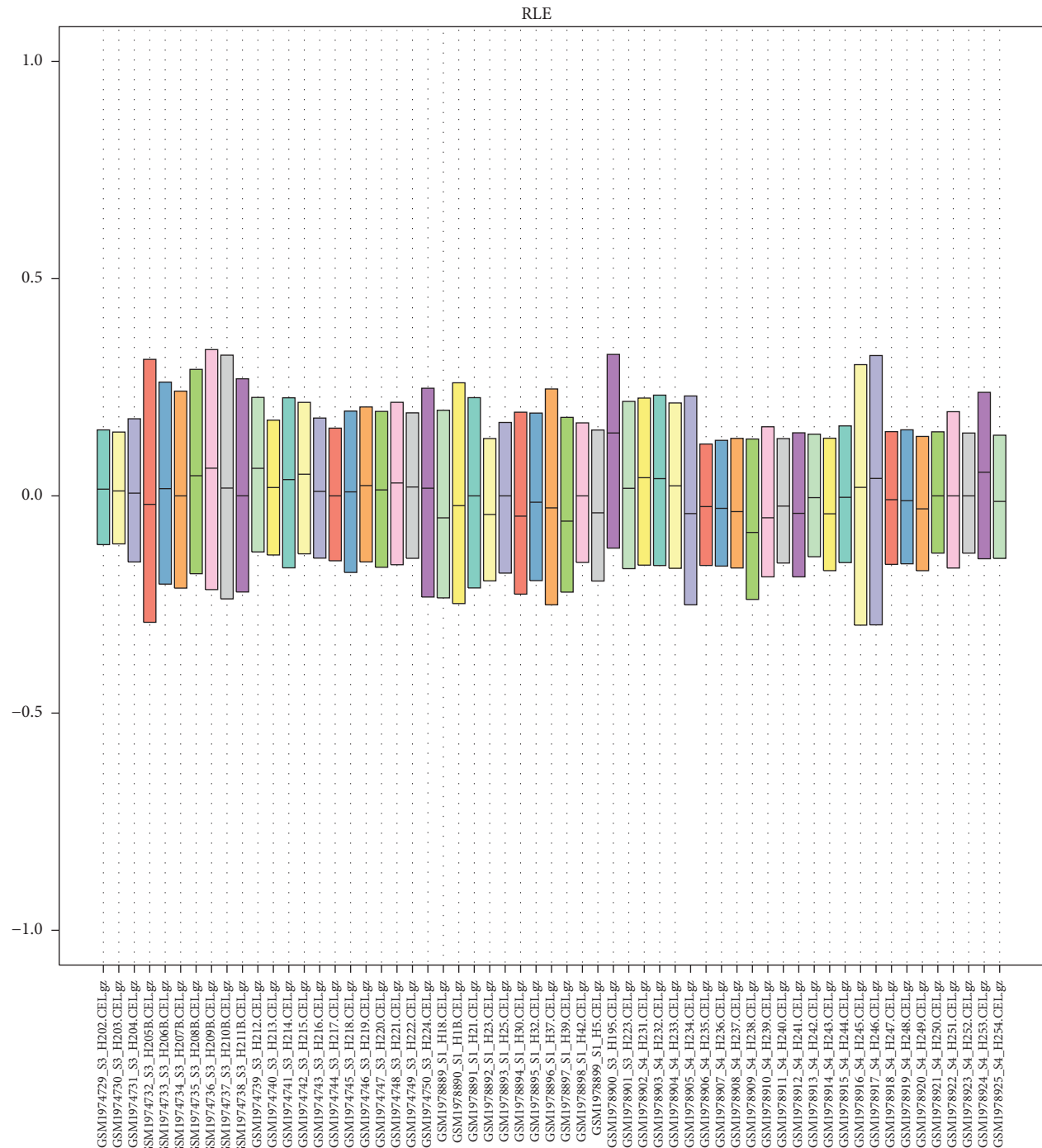
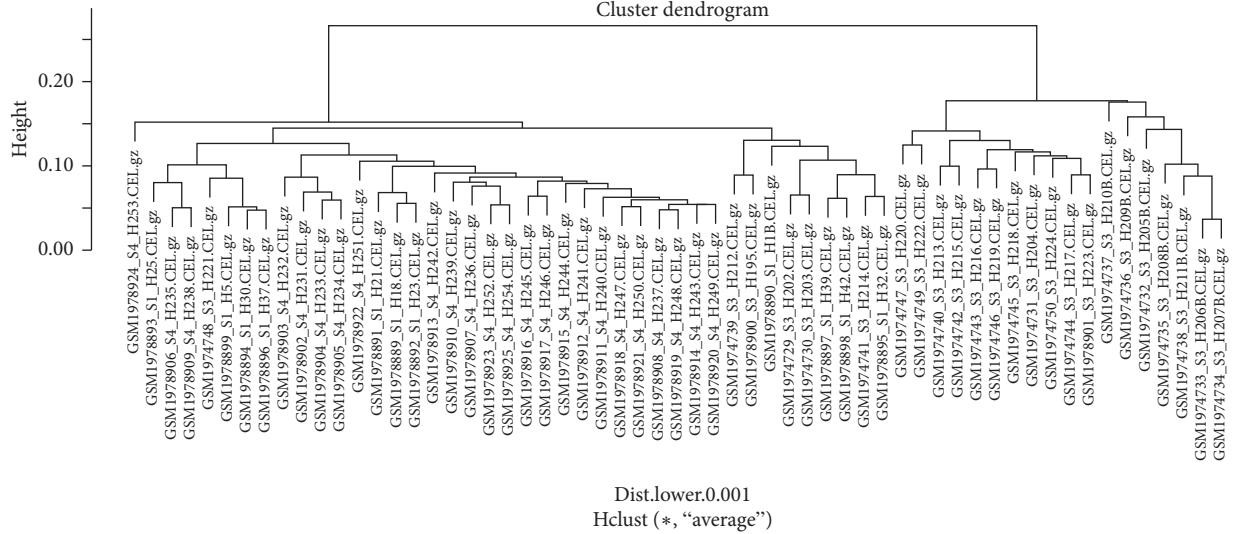


FIGURE 2: RLE box chart.

TABLE 1: Probes number and genes number under different  $P$  values.

|               | $P \leq 0.1$ | $P \leq 0.05$ | $P \leq 0.01$ | $P \leq 0.001$ |
|---------------|--------------|---------------|---------------|----------------|
| Probes number | 24017        | 6251          | 6022          | 5370           |
| Genes number  | 16383        | 5561          | 5361          | 4792           |

FIGURE 3:  $P < 0.0001$  chip cluster diagram.

$$J = \frac{1}{2} \|w\|^2. \quad (8)$$

Calculating that the  $i$ -th feature removal is the change of  $J$ ,

$$\Delta J(i) = \frac{\partial J}{\partial w_i} \Delta w_i + \frac{\partial^2 J}{\partial w_i^2} (\Delta w_i)^2, \quad (9)$$

where  $w_i$  also means that the  $i$ -th feature is deleted; so, as a sorting criterion, the later the feature means that the less information it contains, and the more it will be deleted first. The algorithm is a circular process.

The classic is a linear kernel function with sorting coefficients of

$$\begin{cases} \text{Rank}(i) = (w_i)^2, \\ w = \sum_{i=1}^l a_i y_i x_i. \end{cases} \quad (10)$$

In the case of nonlinearity, it is assumed that in the training sample matrix, when a certain feature is removed, the median of quadratic programming remains unchanged; that is, the obtained classifier does not change. On the premise of this assumption, the contribution value of each feature to the objective function, i.e., the ranking coefficient, is

$$\begin{cases} \text{Rank}(i) = \frac{1}{2} a^T Q a - \frac{1}{2} a^T Q(-i) a, \\ Q_{ij} = K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j). \end{cases} \quad (11)$$

This assumption is also reasonable and feasible in practical application, where  $a = [a_1, a_2, \dots, a_l]$ ,  $Q(-i)$ , means the  $Q$  matrix value calculated when the  $i$ -th feature is assumed to be removed. In practical applications, nonlinear kernels and linear kernels often produce similar results. The SVM-RFE method executes this process iteratively and finally obtains a feature sorting table. Using the sorting list, several nested feature subsets are defined to train SVM, and the advantages and disadvantages of these subsets are evaluated according to the prediction accuracy of SVM, thus obtaining the optimal feature subset. It should be noted that the single feature in the front row does not necessarily make the SVM classifier to obtain the best classification performance, but the combination of multiple features makes the classifier to obtain the best classification performance. Therefore, SVM-RFE algorithm can select complementary feature combinations. The objects targeted by the two formulas are different, corresponding to linear and nonlinear kernels, respectively, but in fact the difference in the final selection of eigenvalues is not obvious.

SVM-RFE algorithm can define a set of nested feature subsets  $F_1 \subset F_2 \subset F_3 \dots F_n$  according to its feature sorting table. The prediction accuracy of SVM is used to evaluate the advantages and disadvantages of these subsets, so as to obtain the optimal feature subset.  $F_i (i = 1, 2, \dots, n)$  means that  $i$ -th features with the highest ranking are selected from the feature set as subsets to ensure that each subset contains features with relatively important information, and then the classifier is designed with the selected optimal subset. The algorithm is as follows in Algorithm 1.



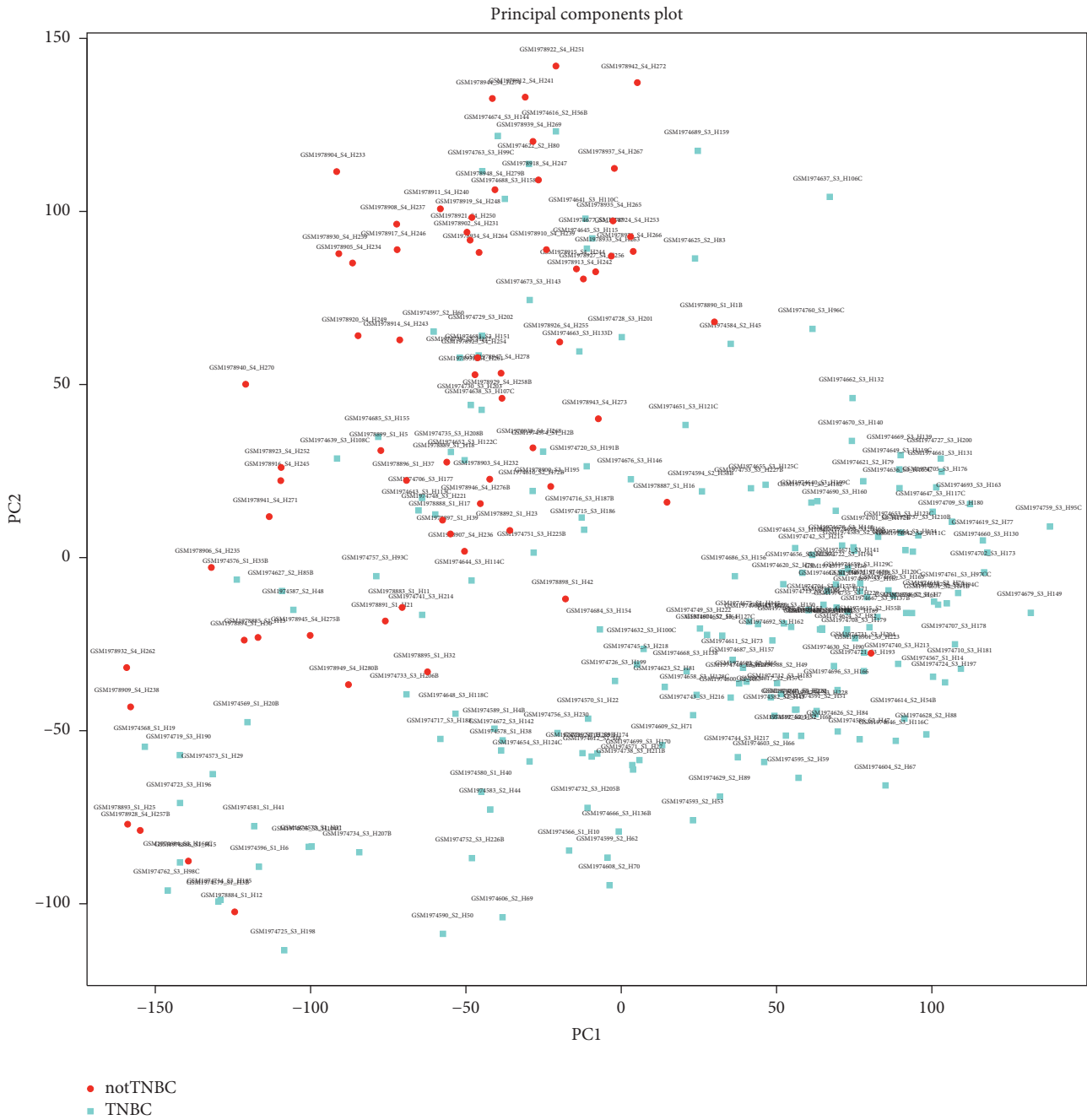


FIGURE 4: PCA differential expression diagram.

### 3. Result Analysis

In this paper, the gene chip GSE76275 is used as the research basis and the relevant parameters of the chip are described. The GSE76275 dataset contains 265 samples, including 198 TNBC and 67 non\_TNBC, with a total of 54613 gene expression values. In the experiment, the relevant basic data are analyzed uniformly, and the expression level of most genes can be kept consistent. The relative logarithmic expression (RLE) box chart can reflect the abovementioned trend. It is defined as the logarithm of the expression value of a probe group in a certain sample divided by the median of

the expression value of the probe group in all samples. The distribution of RLE of all probe groups in a sample can be represented by a box chart commonly used in statistics, and the center of each sample should be very close to the position of ordinate 0. The RLE box chart of this experiment meets this requirement, as shown in Figure 2.

Normalization processing: the purpose of normalization is to enable each group of measurements or measurements under experimental conditions to compare with each other and eliminate nonexperimental differences between measurements, which may come from sample preparation, hybridization process, or

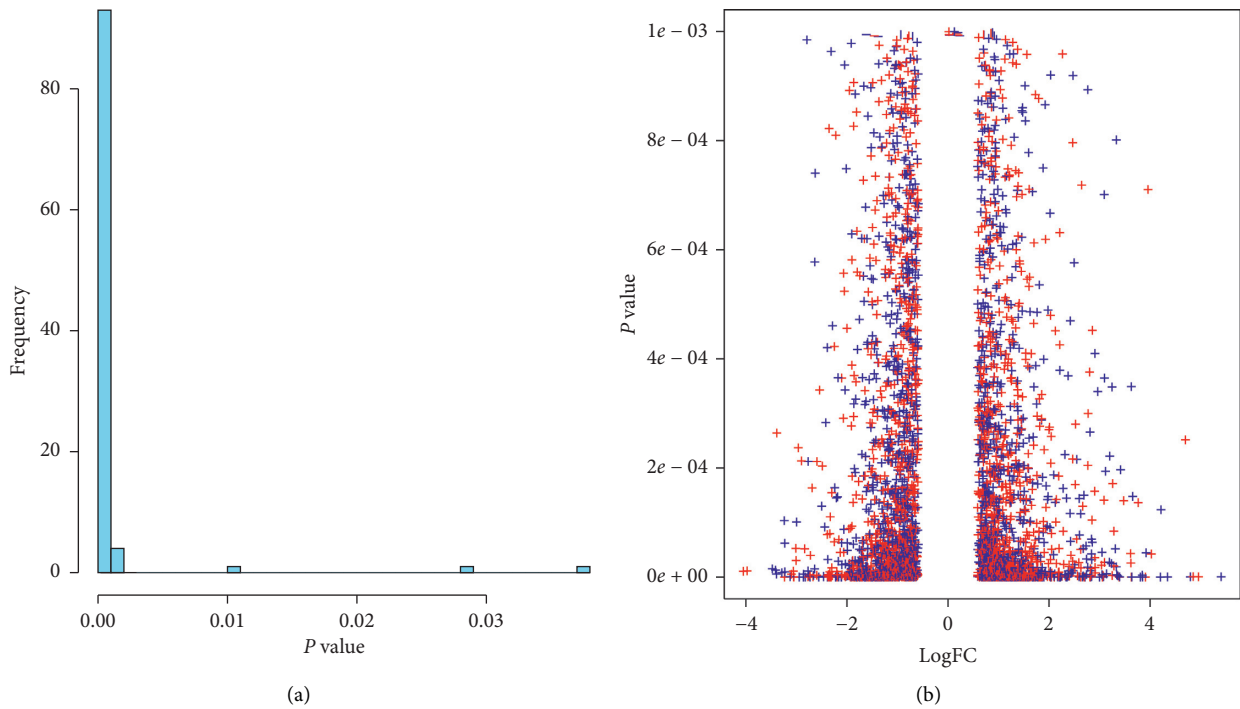


FIGURE 5: The genes number distribution of  $P < 0.001$  and the correlation LogFC and  $P$  value.

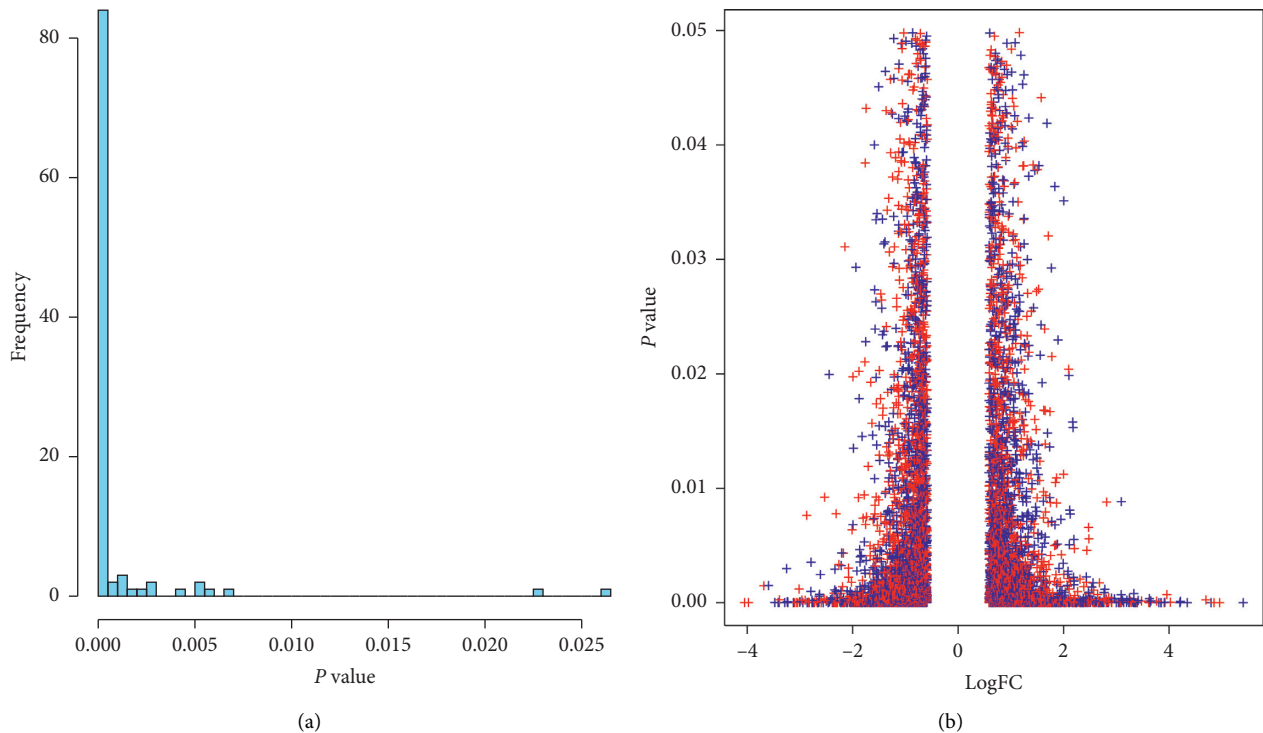


FIGURE 6: The genes number distribution of  $P < 0.05$  and the correlation LogFC and  $P$  value.

hybridization signal processing. The process to normalize the abovementioned data can be implemented by the `expssio` function in the `affy` software package. In fact, the integrated algorithm using preset parameters is more reasonable and efficient.

**3.1. Differentially Expressed Genes Selected.** The first step in the significance analysis of gene expression differences is to select and express genes with significant differences. Generally speaking, the basic assumption of this kind of analysis is that the standardized chip data conform to normal

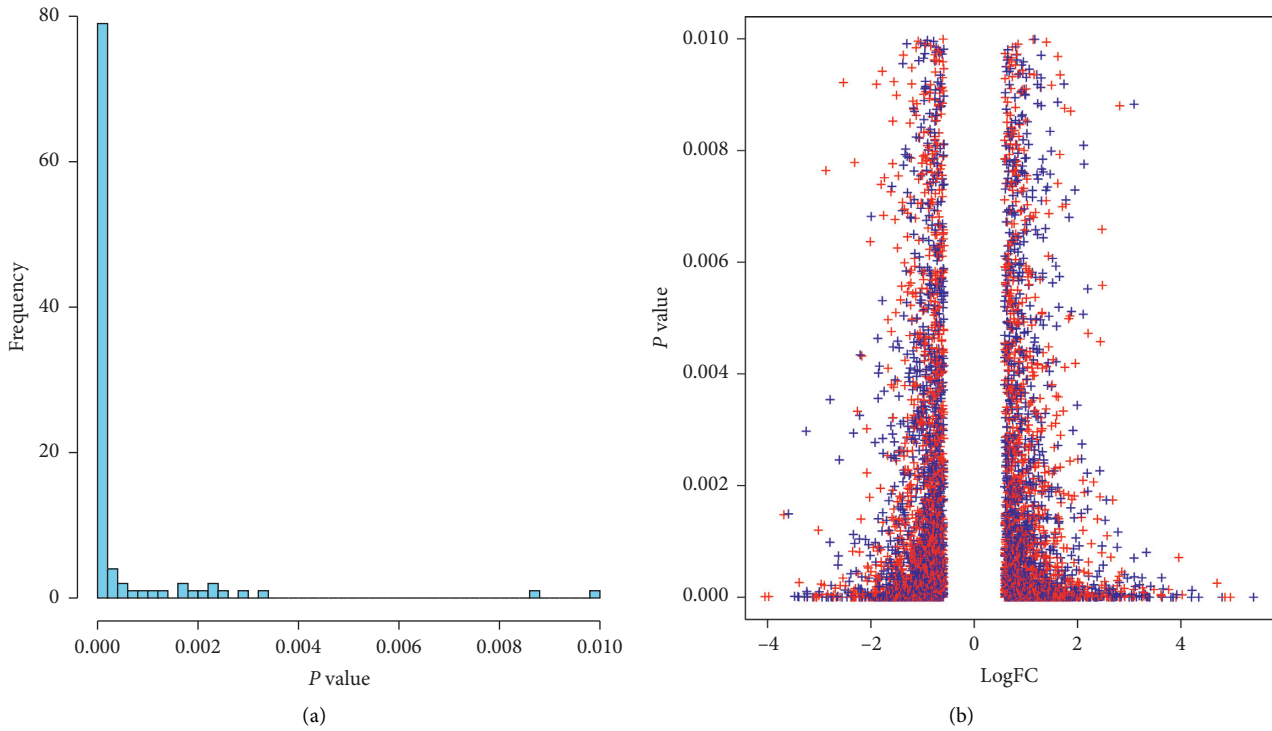


FIGURE 7: The genes number distribution of  $P < 0.01$  and the correlation LogFC and  $P$  value.

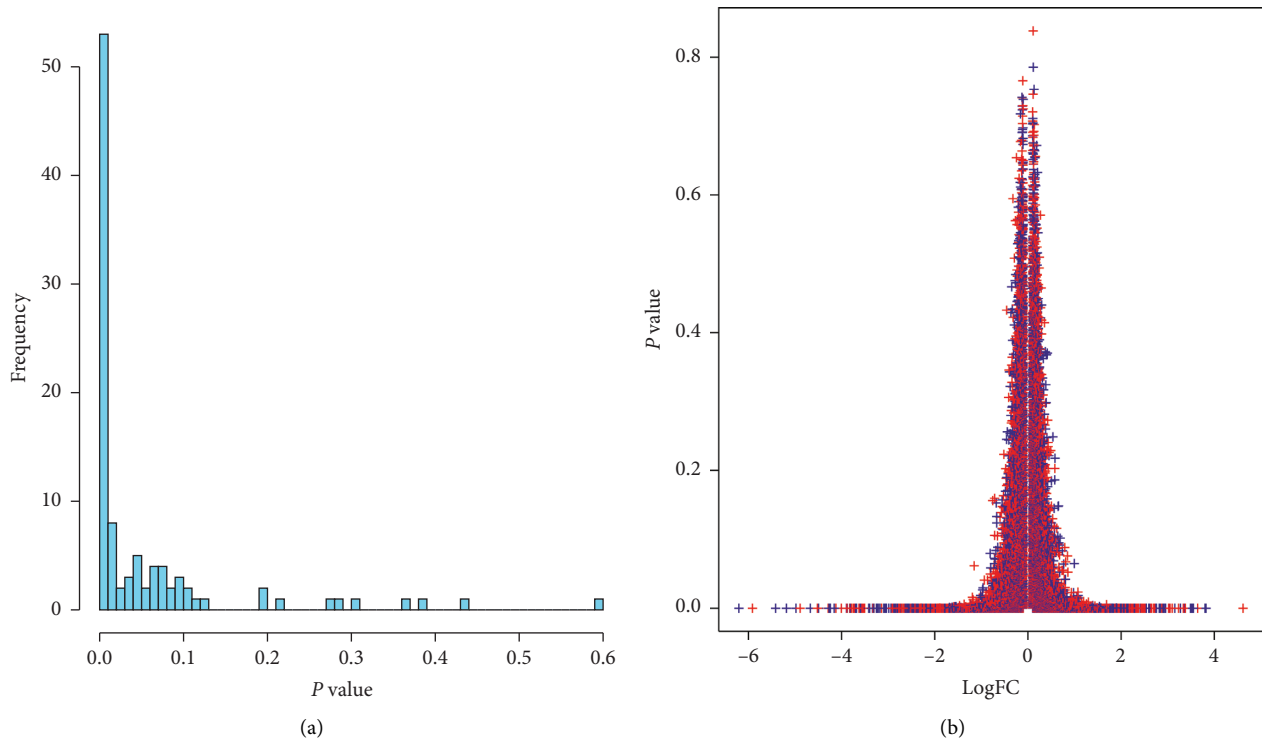


FIGURE 8: The genes number distribution of  $P < 0.1$  and the correlation LogFC and  $P$  value.

distribution, so the statistical methods used are basically  $T$  test,  $F$  test, variance analysis, and the improved forms of these three statistical methods. In order to obtain standardized differential genes, the gene chip adopts the

Bayesian method. Empirical Bayesian method is currently the most commonly used analysis method, which has been completely implemented by limma package of Bioconductor, as shown in Table 1.

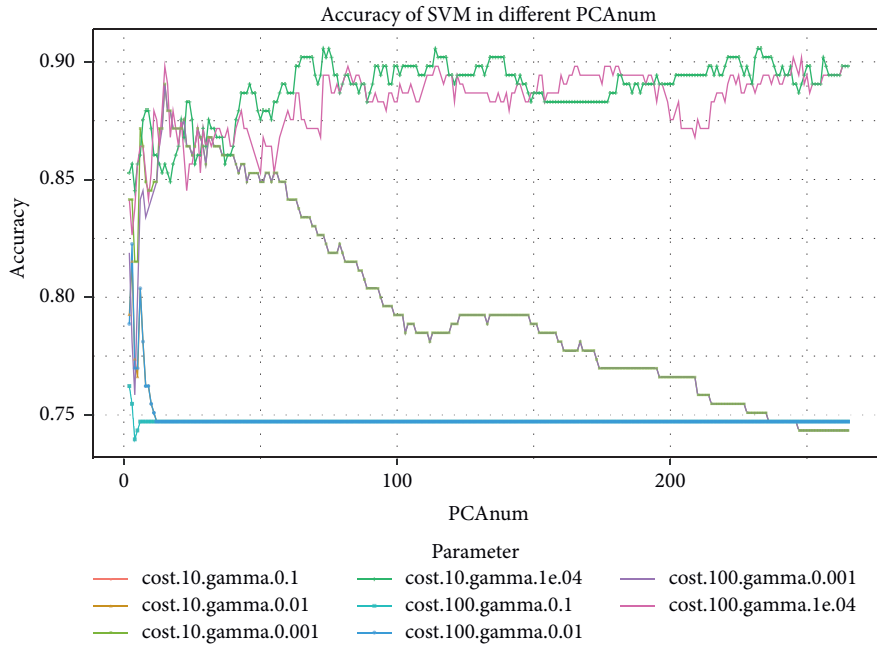


FIGURE 9: Comparison of correct rate of SVM optimization parameters.

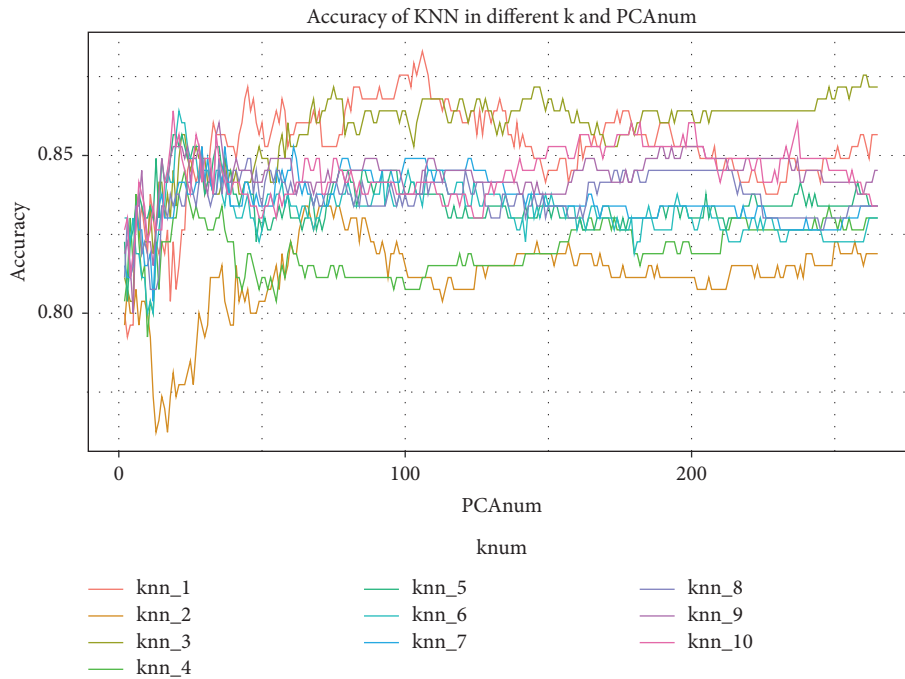


FIGURE 10: Comparison of correct rate of KNN optimization parameters.

The gene clustering of different genes with  $P < 0.001$  was analyzed separately. Some samples are selected for clustering, and the same type of samples can be basically clustered together, as shown in Figure 3.

Select the differential gene expression data of sample  $P < 0.001$  and make a PCA diagram. From the diagram, it can be seen that the classification of the two groups of samples is obvious; thus, it can be seen that the two types of samples have obvious differences, as shown in Figure 4.

When  $P$  is at different values, SVM-RFE shows the difference of screening genes. When  $P$  values are 0.001, 0.05, 0.01 and 0.1, the number and distribution of genes are shown in Figures 5–8:

As can be seen from the above Figures 5–8, when the maximum value of  $P$  becomes larger and larger, the number of genes distributed becomes more and more. However, there is a certain correlation between LogFC and  $P$  value. Most of the points are published between  $[-2, 2]$ , which

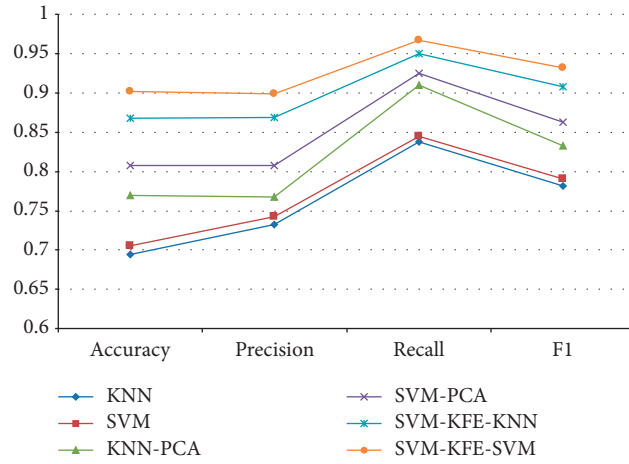


FIGURE 11: The comparison of algorithm performance when  $P \leq 0.05$ .

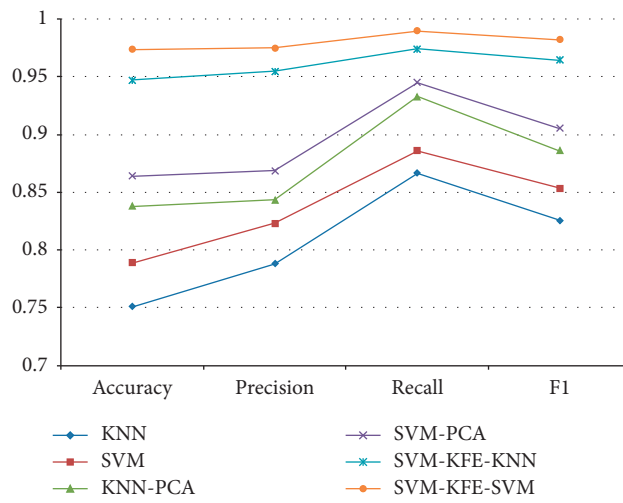


FIGURE 12: The comparison of algorithm performance when  $P \leq 0.01$ .

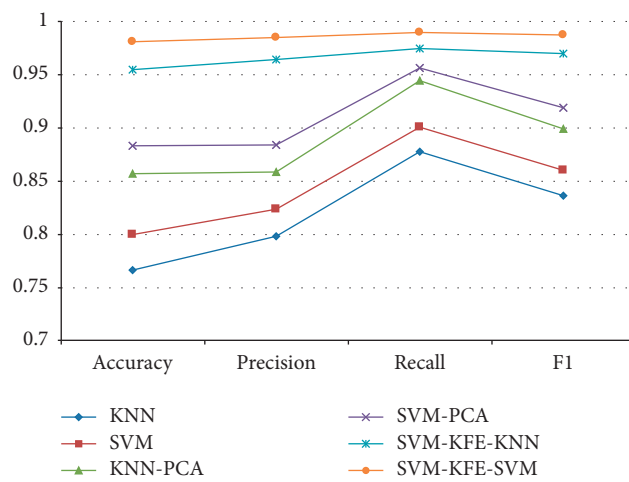


FIGURE 13: The comparison of algorithm performance when  $P \leq 0.001$ .

TABLE 2: Optimal values under SVM parameters.

| Gamma     | Cost | Error     | Dispersion |
|-----------|------|-----------|------------|
| $10^{-6}$ | 10   | 0.1510823 | 0.06924002 |
| $10^{-5}$ | 10   | 0.1512987 | 0.07016349 |
| $10^{-4}$ | 10   | 0.2686147 | 0.07255622 |
| $10^{-3}$ | 10   | 0.2686147 | 0.07255622 |
| $10^{-2}$ | 10   | 0.2686147 | 0.07255622 |
| $10^{-1}$ | 10   | 0.2686147 | 0.07255622 |
| $10^{-6}$ | 100  | 0.1274892 | 0.07413187 |
| $10^{-5}$ | 100  | 0.1512987 | 0.07016349 |
| $10^{-4}$ | 100  | 0.2686147 | 0.07255622 |
| $10^{-3}$ | 100  | 0.2686147 | 0.07255622 |
| $10^{-2}$ | 100  | 0.2686147 | 0.07255622 |
| $10^{-1}$ | 100  | 0.2686147 | 0.07255622 |

accords with the characteristics of normal distribution. The larger the  $P$  value, the larger the number of genes screened out, which shows that selecting the appropriate value of  $P$  for screening can be effectively applied in SVM-RFE algorithm. In order to improve the effect and accuracy of algorithm classification, when  $|\text{LogFC}|$  approaches 0, the larger the range represented by  $P$  value, the more genes there are.

**3.2. Complex Algorithm Parameter Selection.** The parameter selection of the algorithm is an important part of the experiment, and better experimental results can be obtained by selecting better parameters. Therefore, some data in this experiment are selected for experimental parameter selection, and the final experimental comparison is carried out through the selected parameters.

In order to express the best parameter requirements, the cost of SVM is 10 and 100, the gamma is  $10^{-1}$ ,  $10^{-2}$ ,  $10^{-3}$ ,  $10^{-4}$ ,  $10^{-5}$ , and  $10^{-6}$ , and the kernel function is radial. After comparing and optimizing the algorithms, the corresponding error values are obtained under different parameters, and the effects are shown in Table 2.

The distribution comparison of the algorithms shows that when gamma =  $10^{-6}$  and cost = 100, the minimum error value is 0.1274892. There are only 49 samples to build the model, 41 of which are used as support vectors. The proportion of support vectors is too large (over 80%), which indicates that there are irrelevant and redundant features in the used features. It is suggested to use feature selection method RFE to eliminate redundancy and irrelevance and reduce dimension and then use SVM. It is also possible to consider reoptimizing the parameters, but since `tune.svm()` has been used to find parameters, the parameters found are not good. It is better to use fixed parameters and then use RFE for feature screening. Then, this model is used to classify the test set data and use the contingency table to count the accuracy rate, as shown in Figure 9.

When cost = 100 and gamma = 0.0001, cost = 100 and gamma = 0.1, and cost = 10 and gamma = 0.01, the average accuracy is only 75%. When cost = 10 and gamma = 0.001 and cost = 100 and gamma = 0.001, the accuracy of the algorithm is about 88%. With the increase of sample size, the accuracy also decreases. When the sample size is more than 200, the accuracy is less than 75%. When cost = 10 and

gamma = 0.0001, the accuracy of the algorithm is relatively high, about 92%, and relatively stable, as shown in Figure 10.

In the initial stage of the algorithm, the overall accuracy rate is relatively low, only 80%, and the lowest is only 75%. Due to the small sample size, the classification effect is not very ideal. When KNN takes 3, the correct rate is about 90%. When KNN takes 2, the accuracy rate is relatively low, only about 82%. The average accuracy rate of the whole sequencing set is about 84%, and when the sample size increases, the average accuracy rate is relatively stable.

**3.3. Comparison of Algorithms.** When  $P$  takes different values, the differential gene expression data are screened and the selected results are classified. In this paper, several algorithms are selected to screen and analyze genes. The algorithms SVM, KNN, SVM-PCA, KNN-PCA, SVM-KFE-SVM, and SVM-KFE-KNN are used to compare and analyze the performance of accuracy, precision, recall, and F1.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad (12)$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (13)$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (14)$$

In order to evaluate the advantages and disadvantages of different algorithms, the concept of F1 value is proposed on the basis of precision and recall to evaluate precision and recall as a whole. F1 is defined as follows:

$$F_1 = \frac{\text{precision} * \text{recall} * 2}{\text{precision} + \text{recall}}, \quad (15)$$

Through the comparative study of the performance indexes of the abovementioned algorithms, the effects are shown in Figures 11–13.

As can be seen from Figures 11–13, the overall effect of the 6 algorithms is relatively consistent at different  $P$  values. When the  $P$  value selected is smaller, the performance of the six algorithms is improved. In particular, the performance of SVM-RFE-SVM algorithm is obviously improved; accuracy, precision, recall, and F1 are close to 0.99. Among them,

KNN and SVM algorithms have the worst performance because they have no advantages in gene screening effect. SVM-RFE-SVM and SVM-RFE-KNN algorithms have the best results after gene screening and have obvious advantages in gene screening.

#### 4. Conclusion

In this paper, SVM and KNN algorithms are tested, and important indexes such as error rate and accuracy rate of the algorithms are evaluated to obtain the optimal parameters. SVM-RFE-SVM was proved to be effective by screening and comparing SVM, KNN, KNN-PCA, SVM-PCA, SVM-RFE-SVM, and SVM-RFE-KNN binding genes. In the later research work, the effectiveness of the algorithm proposed in this paper is tested in different datasets, and normalization is carried out in unbalanced datasets for classification research. The effectiveness of the classification algorithm is analyzed by combining the number of exons and mutations of gene sequencing data in RNA-SEQ. Correlation analysis between different types of sequencing data is the ultimate goal of the research work.

#### Data Availability

The data used in this research are available in the website <https://pan.baidu.com/s/1e3du8VbzjnvxHRhunF1o0Q> (download code: yzjz).

#### Conflicts of Interest

The authors declare that they have no conflicts of interest.

#### Acknowledgments

This work was supported by the National Nature Science Foundation of China under contract numbers 61271259 and 61601070; the Chongqing Nature Science Foundation under contract numbers CTSC2011jjA40006, CSTC2010BB2415, and CSTC2016jcyjA0455; the Research Project of Chongqing Education Commission under contract numbers KJ120501, KJ12050, KJ1600411, and KJ110530; the Key Project of Science and Technology Research of Chongqing Education Commission (KJZD-K201800603 and KJZD-M201900602); the Chongqing Graduate Scientific Research Innovation Project under Grant no. CYB17131; and the Doctoral Candidate Innovative Talent Project of Chongqing University of Posts and Telecommunications under Grant no. BYJS2016003.

#### References

- [1] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, pp. 389–422, 2002.
- [2] K. B. Duan, J. C. Rajapakse, and M. N. Nguyen, "One-versus-one and one-versus-all multiclass SVM-RFE for gene selection in cancer classification," in *Proceedings of the Evolutionary Computation Machine Learning and Data Mining in Bioinformatics, European Conference, Evobio 2007*, pp. 47–56, Valencia, Spain, April 2007.
- [3] Y. Ding and D. Wilkins, "Improving the performance of SVM-RFE to select genes in microarray data," *BMC Bioinformatics*, vol. 7, no. S2, 2006.
- [4] S. Yoon and S. Kim, "Mutual information-based SVM-RFE for diagnostic classification of digitized mammograms," *Pattern Recognition Letters*, vol. 30, no. 16, pp. 1489–1495, 2009.
- [5] Y. Tang, Y.-Q. Zhang, and Z. Huang, "Development of two-stage SVM-RFE gene selection strategy for microarray expression data analysis," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 4, no. 3, pp. 365–381, 2007.
- [6] X. Huang, B. L. Zhang, and Z. ZhangLi, "Feature clustering based support vector machine recursive feature elimination for gene selection," *Applied Intelligence*, vol. 48, no. 3, pp. 594–607, 2018.
- [7] J. Yin, J. Hou, and Z. She, "Improving the performance of SVM-RFE on classification of pancreatic cancer data," in *Proceedings of the IEEE International Conference on Industrial Technology (ICIT)*, pp. 956–961, IEEE, Taipei, Taiwan, May 2016.
- [8] R. Wang, R. Li, and Y. Lei, "Tuning to optimize SVM approach for assisting ovarian cancer diagnosis with photoacoustic imaging," *Bio-Medical Materials and Engineering*, vol. 26, no. s1, pp. S975–S981, 2015.
- [9] C. Chen and H. D. Zhu, "Feature selection method based on parallel binary immune quantum-behaved particle swarm optimization," *Advanced Materials Research*, vol. 546–547, pp. 1538–1543, 2012.
- [10] A. Anaissi, M. Goyal, D. R. Catchpoole, A. Kennedy, and P. J. Braytee, "Ensemble feature learning of genomic data using support vector machine," *PLoS One*, vol. 11, no. 6, Article ID e0157330, 2016.
- [11] V. Bolón-Canedo, N. Sánchez-Marroño, A. Alonso-Betanzos, J. M. Benítez, and F. Herrera, "A review of microarray datasets and applied feature selection methods," *Information Sciences an International Journal*, vol. 282, no. 5, pp. 111–135, 2014.
- [12] T.-C. Chen, Y.-C. Hsieh, P.-S. You, and Y.-C. Lee, "Feature selection and classification by using grid computing based evolutionary approach for the microarray data," in *Proceedings of the 3rd International Conference on Computer Science and Information Technology (ICCSIT)*, pp. 85–89, IEEE, Chengdu, China, July 2010.
- [13] X. Zhou and J. Wang, "Feature selection for image classification based on a new ranking criterion," *Journal of Computer and Communications*, vol. 3, no. 3, pp. 74–79, 2015.
- [14] L. Zhang and X. Huang, "Multiple SVM-RFE for multi-class gene selection on DNA microarray data," in *Proceedings of International Joint Conference on Neural Networks (IJCNN)*, pp. 1–6, IEEE, Killarney, Ireland, July 2015.

## Research Article

# Automatic Grading for Complex Multifile Programs

Tiantian Wang <sup>1</sup>, Djoko Budi Santoso,<sup>1</sup> Kechao Wang,<sup>2</sup> and Xiaohong Su<sup>1</sup>

<sup>1</sup>*School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China*

<sup>2</sup>*School of Information Engineering, Harbin University, Harbin 150001, China*

Correspondence should be addressed to Tiantian Wang; wangtiantian@hit.edu.cn

Received 31 March 2020; Accepted 23 May 2020; Published 13 June 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Tiantian Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents an automatic grading method DGRADER, which handles complex multifile programs. Both the dynamic and the static grading support multifile program analysis. So, it can be an advantage to handle complex programming problem which requires more than one program file. Dynamic analysis takes advantage of object file linker in compilation to link complex multifile program. The static grading module consists of the following steps. Firstly, the program is parsed into abstract syntax tree, which is mapped into abstract syntax tree data map. Then, the information of preprocessor is used for linking external sources called in main program by complex multifile program linker-fusion algorithm. Next, standardization process is performed for problematic code removal, unused function removal, and function sequence ordering based on function call. Finally, program matching successfully tackles structure variance problem by previous standardization process and by simple tree matching using tag classifier. The novelty of the approach is that it handles complex multifile program analysis with flexible grading with consideration of modularity and big scale of programming problem complexity. The results have shown improvement in grading precision which gives reliable grading score delivered with intuitive system.

## 1. Introduction

Automatic Grading System (AGS) is program which can determine student grade automatically based on score objective parameter. AGS is needed as popular courses especially in computer science often have hundreds or thousands of students but only a few staff [1]. Programming assignment in these courses is necessary to improve technical programming and problem solving skill of students. Manual assessment method of programming practices is a tedious and time-consuming task [2, 3]. This method is inadequate because first programming course are often typically complex for a lecturer to assess correctness accurately and comprehensively by manual assessment [4]. In here, AGS can become a key role to maintain accuracy and avoid biases as the grading process is based on objective scoring rules. Importantly, the system must provide immediate feedback to students so they can learn from their mistakes [5], allowing them to make self-learning without an instructor [6].

With the development of information technology, more and more schools and organizations try to realize the

inclusive and fair education through online learning, human-computer interaction [7, 8] learning, etc. Nowadays, various AGS systems already exist and are used by higher education institutions to enhance learning process [9, 10] or as programming communities to self-improve in programming and problem solving skills. The question “why do so many automatic assessment systems exist, and why are new ones created every year?” pops up into discussion in this research field [11]. One of the reasons for various AGS is that each system may deliver some features which are not provided by others to distinguish their advantage factor.

The research still continues until today to pursue reliable grading system and solve the issues existed in this field. In here, current major issues are listed, which work as a sample. This gives an idea to produce final reliable grading and potential future work.

According to Table 1, these are recent knowledge issues chosen as the main objective overview in this work to do novelty works.

Firstly, code variance which can be said as the biggest factor to affect the result of static program analysis. In static



TABLE 1: Current research issues in automatic grading system.

| Issue   | Approach   | Results   |
|---|--|---|
| Code variations are widely believed to impede program analysis because various source codes have to be recognized as the same [12]  | Program normalization by using system dependence graph (SDG) implemented in prototype called normalizer  | Successful for small-sized program, not big-sized or complex program                              |
| Most existing inlining algorithms not suitable for code analysis and interprocedural analysis can analyze the calling context but are very costly and cannot remove function call variations [13] | Inlining algorithm based on program dependence graph (PDGs) using simple function call tree (SCFT) applied to code normalization                           | Limited in dealing with programs with multiple files  |
| Deals with multiplicity of solutions that exists for the same programming problems providing automated evaluation: immediate feedback for students and grading assignment of program [14]         | Partitioned block by using control flow graph (CFG) and program behavior comparison with symbolic execution rather than textual context or concrete values | Solves a multitude of solution of problems but is complex in implementation for big scale program |

analysis, these issues must be considered to produce reliable and precise grading score. Code variance issues solved by introduced program normalization mechanism to change the program into an intended structure requiring a program representation. The program normalization approach uses system dependence graph as the program representation. It requires for changing the code or program component known as refactoring process. The results of the works that solve code variance issue have significant impact of grading quality result.

Secondly, state that current inlining algorithms are not suitable for code analysis. The work approach by using program representation in Program Dependence Graphs to expand and manipulate program component focused on function call is introduced. This issue is actually related to code variance and focused on the structure program, which is influenced by function call and function sequence order written in program code. Results of the approach give more suitable inlining algorithm for code analysis and improve mechanism by reducing step of conventional interprocedural analysis. However, the existing work cannot deal with functions in several files.

Thirdly, Arifi et al. have introduced how the program is partitioned by a block using Control Flow Graph and generated symbolic variance for program comparison to solve the problem of multiplicity of solutions [14]. This work solves a multitude of solution of problems, but is complex in implementation for big scale programs.

These research studies lay a good foundation for our work. However, further work is still needed to support complex multifile programs. As the current single file program analysis is a relatively straight forward and common, complex multifile program analysis is a plus to handle big complex program which is still not solved or suitable for the recent research [15]. In this paper, we present a novelty work to handle complex multifile program analysis.

## 2. Related Work

*2.1. Complex Multifile Program Analysis.* AGS should be capable to quickly grade multiple and complex computer literacy assignments while providing meaningful feedback in order to stimulate an efficient learning process [16].

Specially, for larger and complex programming assignment, it is still not possible for most existing systems to assess good programming solution. In this case, semiautomated system is still used, which requires the human evaluator to use partial part of results from system to decide final grading [17]. It makes the task more complicated for humans to assess modular code. This issue will be faced by the current grading system as a programming problem is more scaled and complex for advance programming courses or real practice programming project. It makes the automatic grading system not follow their primary principle as an automatic system and become ineffective.

Complex multifile program analysis can be an advantage to handle complex programming and scaled project, which requires more than one file program. As stated before, it also makes modularity and flexibility such as creating user-based custom libraries to support core program file without rewriting into main program. In this feature, the implementation of complex multifile program analysis uses their primary concept.

*2.2. Dynamic and Static Grading Methods.* Dynamic analysis uses black-box concept which depends on output results. The analysis requires program to be compiled and run with test cases. The final grade uses comparison results of the produced output with the expected output. However, it has fatal drawback, in which student program may not produce an output because problematic code such as syntactical errors makes the program fail in the compilation process [18]. This means the program will fail before processing test cases. Hence, dynamic analysis itself is not enough for completely giving all objective scores when it fails and static analysis is needed.

Static analysis uses white-box concept without compiling and running the program. It uses rapid advancement technology from compiler and language-based tool. The approach also uses code analysis knowledge to analyze program and gives grading based on scoring objective rules and parameters.

However, both methods are having their advantage and disadvantage since static analysis require more complex process of code analysis. It requires predefined rules and

objectives. On the contrary, dynamic analysis is a more direct approach but does not cover all aspect, especially, when it fails to perform grading. Choosing which methodology is more feasible makes dynamic versus static analysis become a topic for consideration for the grading system. According to dynamic and static analysis principle in the previous section, both methods have their advantage and disadvantage factors. Table 2 summarizes of their comparison.

In here, the knowledge summarizes that static analysis cannot be used for checking the correctness of student programs using test cases as an input which produces an output. On the contrary, traditional dynamic analysis systems will completely fail to perform grading and miss important aspects when assessing student programs such as checking the code quality [19]. This may be the reason to why some existing automatic grading systems combine the best of both approaches by improving dynamic testing mechanism with static technique. By carrying out mutual combination of both analysis and providing immediate feedback in grading result, it gives an additional positive value to the user and advantage of the grading system.

Back to 1992, when Cellidh [20] was introduced, it was actually a pioneer grading system which combined both approaches by introducing semantic error detection. It used for detecting infinite loop issue which is critical for dynamic grading. The system also uses static verification including structure, indentation, detect comment, readability measurement, and complexity metrics in dynamic analysis.

In 1997, system called ASSYST [21] combined both analyses in practice to automate some aspects of grading for introductory Ada classes, as well as a second-year C programming course. It gives grading score to students based on the correctness (actual output compared to the expected output), efficiency (run time) on dynamic analysis side, check program source code style, and its complexity on static analysis side. In 2000, more systems adopt this combined method for providing flexible analysis and pursuing potential grading system which provide meaningful intermediate feedback.

In 2006, Marmoset [22] was built in the University of Maryland. The main purpose of the system is to collect information about development of student programming skill while doing a programming assignment for triggering self-improvement. The unique feature is allowing a full snapshot about student progress in the system, so it can be analyzed in detail by using different types of test cases (student, public, release, and secrete) and a personal support from the lecturer through comments on the code in the page.

In 2008, Web-CAT [23] provided extensibility and flexibility as its plugins-based architecture taking advantage of recent development technologies was built using Java servlet. It provides security features by authentication, erroneous or dangerous code detection, and portability. It also supports manual grading by allowing the lecturer to check program submitted by the student. It allows lecturers to give the comment, suggestion, and grading modification. The programming language supported are C or C++, Java, Pascal, Prolog, and others flexibility support for integration.

The grade is based on correctness through test cases, completeness of program, and validity.

In 2011, eGrader [24] provided detailed feedback reports and allowed students to see model solution provided by the lecturer or course owner. It also gives specific comments on syntax and semantic errors if occurred. The static analysis process implemented in the system consists of two parts, which are the structural similarity and quality analysis. Structural similarity analysis is based on the graph representation of the program. Quality analysis was achieved by measurements using software metrics.

In the same year, a system called AutoLEP [25], as an automatic grading system tool, was developed. It improves the traditional static grading mechanisms by combining dynamic code testing. The approach is enriching static analysis in source code analysis with a comparison of the similarity degree of compared program. The dynamic analysis was used to evaluate correctness of the submitted program using test cases and comparing the expected output. The static analysis does not compile or execute the programs. It uses model program to evaluate student program construction and how close the student source code is from the correct solution which is model programs provided in programming assignment by the lecturer. The final grading result was achieved by calculation of summarization from each grading analysis. The works were reported to distinguish syntactic and semantic analysis from the previous work. The architecture includes (1) the client and a computer used by a student, and it performs the static analysis and can provide quick feedback; (2) a testing server which has to perform the dynamic analysis; and (3) a main server which has to control the information of the other components to establish a grade.

In 2012, a new automatic grading system called Quimera [26] was built as a web-based application. It was able to evaluate the program source code written in C language and provide a full management system for programming contests. It also allows to create and manage programming exercises both in competitive learning and programming contest environments. Besides the traditional dynamic approach, this system provides a static analysis of the program by measuring the source code quality. Thus, the final grade is based not only on the source code capability of producing the expected output but also on its quality and accuracy.

Finally, these are examples of automatic grading tools existed with flexible code analysis which primarily combine both method to achieve advantages of grading in their grading system. The listed tools will be used in comparison analysis.

### 3. Overview

As mentioned before, our approach is implemented in our automatic grading system tools called DGRADER as a web-based online automatic programming judgement platform. Web platform chosen as a researcher focused on web-based assessment system shows positive influence on learning effectiveness [27]. It is also very effective in distributing material and collecting the student assignment online [28].

TABLE 2: Dynamic versus static analysis in the automatic grading system.

| Method  | Concept   | Advantage  | Disadvantage  |
|---------|---|--|---|
| Dynamic | Black-box (output) with test cases                            | (i) More direct for grading<br>(ii) Correctness checking<br>(iii) Popular usage  | (i) Requires compilation process and fails when compilation is unsuccessful<br>(ii) Security issues<br>(iii) Does not cover all aspect of grading |
| Static  | White-box (code analysis) and comparison of the correct model | (i) Does not require compilation process<br>(ii) Capable to analyze code quality | (i) Computational complexity<br>(ii) Requires set of rule definitions for giving reliability grading<br>(iii) Not providing correctness checking  |

Our tool is built by using Spring Model View Controller (MVC) as framework foundation with Spring Tool Suite (STS) Integrated Development Environment (IDE) written in Java programming language. It uses several technologies integrated to the system to support MOOCs. The full system will be running on the host server which can support Apache Tomcat Server. The database technology used in this application is MySQL database.

Architecture: Figure 1 shows three blocks of the system, i.e., user, application (front and back-end), and expanded core grading API block. Every core page contains a block of the modular page with defined web services and API. The system manages basic web features provided by Spring such as servlet, session handler, and its core features.

Core grading Application Programming Interface (API) of the system is shown on the right side. It has several API for specific purposes and functions. The API is triggered by using defined query parameter and handled by the API handler. Core grading APIs cover main functions of DGRADER for the grading task. The first layered API is the assignment handler. Its main purpose is to handle raw materials of assignments and extract the information before the grading task. Submission API is the connector triggered by submission activities. Its role is to forward information to grading analysis process in the grading analysis APIs. The information passed by the query parameter is used to target specific grading in the system. Each grading will be detailed in its section. Finally, the grading activities will produce an output feedback which will be displayed in the interface to users.

**3.1. Main Features.** The main features of DGRADER are as follows:

- (1) Providing e-learning environment management for programming courses.
- (2) Supporting flexible grading analysis: dynamic and static analysis. Flexible grading makes the instructor or programming assignment task creator become more flexible to choose assessment methods. The grading method can be based on the complexity of programming task or its purpose. It also makes the system capable to cover grading assessment task when one of the grading fails.
- (3) Supporting complex multfile program analysis in submission activity for solving one programming problem.

- (4) Providing instant feedback and result of grading analysis for programming assignment problem. The grading is provided in both categorical and numerical results.

Feedback results in static analysis include

- (i) Presenting the final linked-fused program source code and its standardization result
- (ii) Presenting AST traverse log with visitor activities
- (iii) Providing visualization for transformation graph comparison for final linked-fused and standardized program structure
- (iv) Presenting function information such as function list in program, function call sequence, and unused function list removed by standardization
- (v) Presenting the original AST data mapper of final program and standardized AST data mapper of standardized program
- (vi) Providing root tree visualization by AST root data mapper
- (vii) Providing experimental AST visualization of standardized program by standardized AST data mapper
- (viii) Presenting AST simple tree matching trace analysis in data mapper
- (5) Flexible source program uploading or submission by using file upload or directly using integrated CodeMirror text editor.
- (6) Supporting course system management similar to Massive Open Online Courses (MOOCs) and Learning Management System.

## 4. Multiple Program Files Analysis

Multiple program file analysis is a novel feature of our automatic grading system. The reason to present this feature is because existing systems usually support single file to keep simplicity of its grading analysis process. However, if a programming problem is more complex and users need more files, this will become a limitation factor. In other cases, users may want to use some libraries which do not require to be written again or the compiler in dynamic analysis does not support the library. In order to support multiple program file submission, this section presents implemented approaches including file model, multiple program file linker in each grading process, and linker-fusion algorithm.

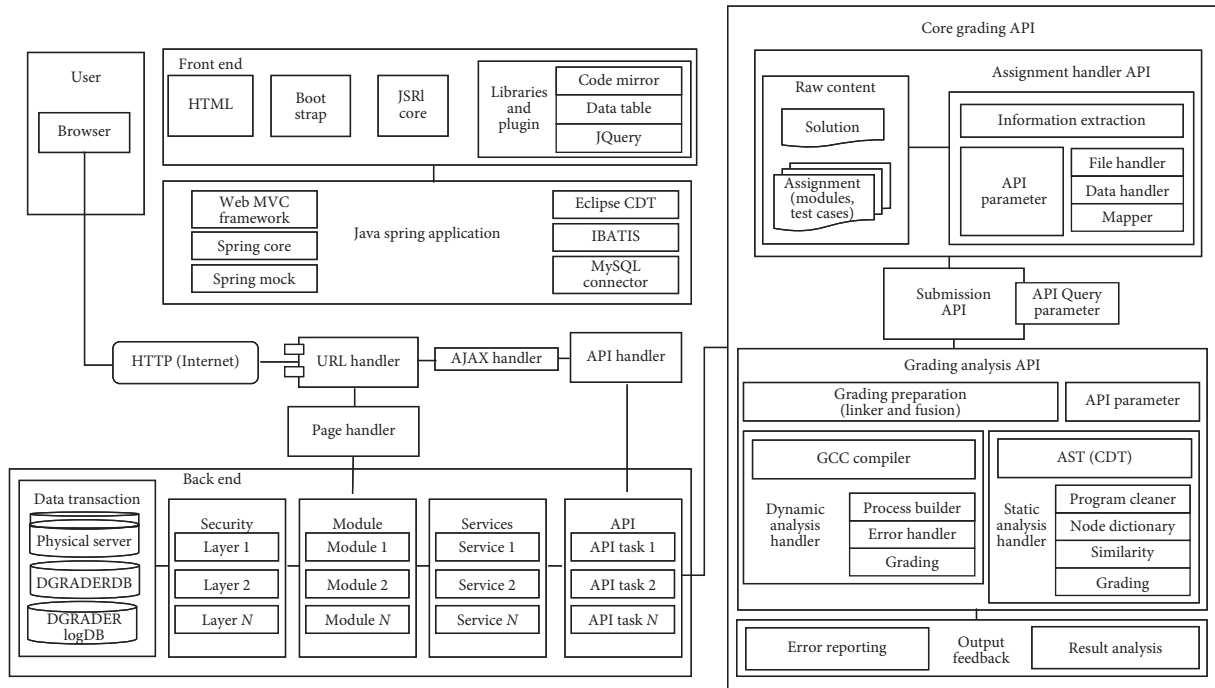


FIGURE 1: System architecture.

4.1. *File Model.* File model is implemented to handle multiple file program submission. It separates into the following two parts:

- (1) **Preprocessor:** the beginning part of source statement or preprocessing statement `#include<[library]>` which can be parsed by using `ASTPreprocessorStatement`. The user can use several external files to support main files for solving one programming problem. The filename of the sources should be the same in file upload process because it will be linked via parameter search. Custom libraries by the user can be written in this defined statement `#include "[external_sources_filename]"` as a rule. The double quote indicates its user custom libraries or external sources. The system will process linker-fusion process to combine linked multiple program file.
- (2) **Content:** this part contains the body of source program or other parts below preprocessor statements.

4.2. *Multifile Linker in Dynamic Grading.* Dynamic grading use compilation for handling complex multfiles to be linked, as shown in Figure 2. As an example for linking complex multifile program C or C++ programming language, it uses the linking process of GCC compilation process. It uses two input: main source and path of external source (multiple files paths). The differentiation from static is the type of the main source in here which will become .bin files from `ProcessBuilder`. The linking process in the system is possible with GCC command by using the parameter in backend program as an example:

```
> gcc-o [main].exe [multiple_files_path] -w (for C
complex multifile program) > g++ -o [main].exe [multi-
ple_files_path]-w (for C++ complex multifile program)
```

`Multiple_files_path` expression is files' path of external sources used in `main_source`. The command will process the main program in the compilation process to become an object and call every external source in the linking process. The linking process will continue to analyze and expand the preprocessor to make every external source become intended object of executable file with `-o` command. The `-w` command used to give warning error feedback such as preprocessor is not linked in case of file not found which makes compilation process fail.

4.3. *Multifile Linker in Static Grading.* The process still has two inputs: main source (raw source) and the external source path of multiple files which are already generated in the preparation process. The two inputs will be linked and fused in the next process, as shown in Figure 3 by linker and fusion. Static grading using source program linker-fusion algorithm is in Algorithm 1. Final output is a fused program with all the source code as a whole. The fused source code will not be processed through compilation but processed with code analysis using AST to parse element of program in external source used in the main program.

The algorithm is divided into four steps:

Step 1: get preprocessor statement set  $ps$  of the main source program by Eclipse CDT/JDT API features which can generate AST of main source using `getAllPreprocessorStatements()`.

Step 2: handling preprocessor statement set  $ps$ . Every  $ps$  found will be filtered by using `patternMatch()` to detect

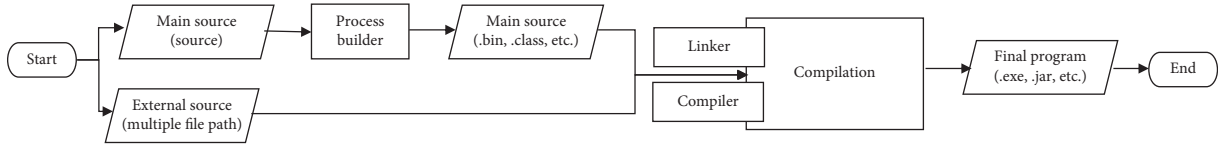
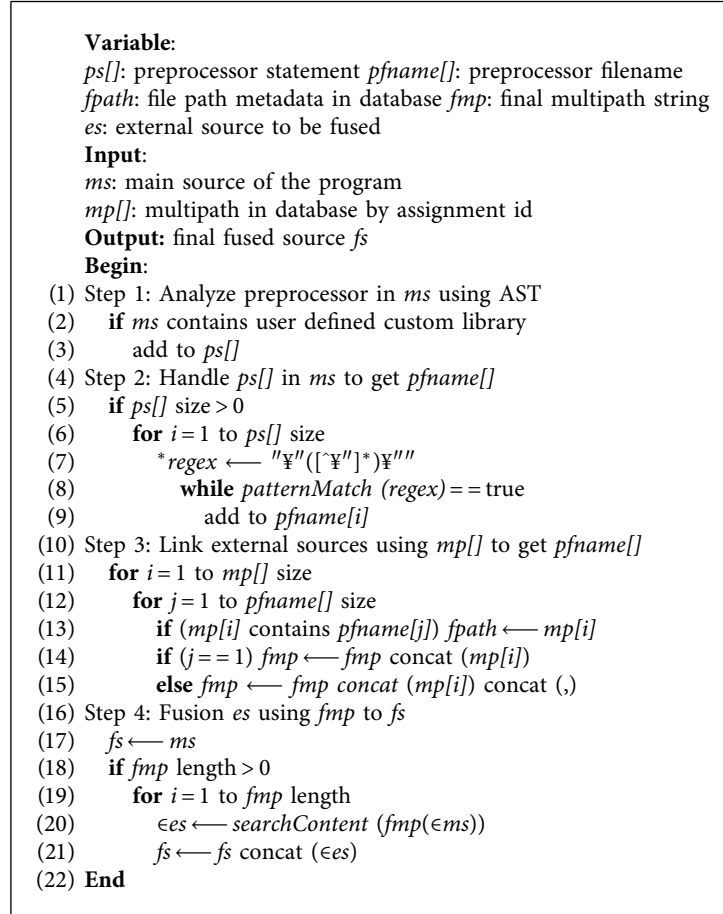


FIGURE 2: Complex multifile program linker in dynamic grading.



ALGORITHM 1: Multifile program linker and fusion algorithm.

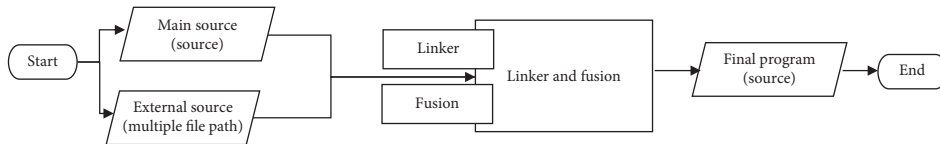


FIGURE 3: Multifiles linker in static grading.

user custom libraries defined by the following format:  
`[TYPE_OF_LANGUAGE_IMPORT_WAY(import/include)]<space>["FILENAME"].` Regex is used to parse preprocessor filenames into  $pfname$  set.

Step 3: linking process by analyzing the file path database  $mp$ . Each  $pfname$  found  $mp$  will generate  $fmp$  as final linked string query for every external source used in main source of the program.

Step 4: final step of the process is to get fused source code  $fs$ . The  $searchContent()$  function will analyze the

code by using textual search (e.g., `.dll.h`) or AST (source code which can be parsed) to get the contents of the files, denoted by  $\in es$ , used by the main source of the program. These contents are fused into one source file.

**4.4. Dynamic Grading.** The dynamic grading method covers multiprogram file submission for solving one programming assignment. The main source file triggered in here stated as a student program with several preprocessor statements is used to direct the external sources files. The first phase is

**Variable:**

ASTdata[]: program AST data map

fASTdata[]: final refactored program AST data map

**Input:**

pAST: program AST

sb[]: syntax bank

**Output:** final standardized program AST fsAST**Begin:**

- (1) Step 1: Traverse and visit pAST with ASTVisitor class, do Step 2
- (2) Step 2: Handle AST node for AST data mapper to get ASTdata[]
- (3) Index = 0
- (4) for every AST node
- (5) initialization ASTdata[] key (see detail in Table 3)
- (6) node = AST node, current tag = AST tag
- (7) if current tag = function statement or expression
- (8) syntax = nodeParser(node)
- (9) if syntax not exists on sb[]//syntax classifier
- (10) inner user function node
- (11) else
- (12) inner standard function node
- (13) if current tag  $\neq$  previous tag
- (14) exit no = index (exit AST node branch tree)
- (15) node info = flag based on node and tag information, process no = index
- (16) index ++
- (17) add ASTdata[index]
- (18) Step 3: Handle AST data for refactoring to get final ASTdata[]
- (19) fASTdata[] = RefactoringCore(ASTdata[i])
- (20) Step 4: Rebuild final source from fASTdata[] to get fsAST
- (21) for  $i = 0$  to fASTdata[] size
- (22) fsAST = fsAST add node(fASTdata[i])
- (23) **End**

ALGORITHM 2: Program standardization using AST algorithm.

TABLE 3: AST data map.

| Key               | Description  |
|-------------------|--|
| Index             | Used for indexing of visiting and traversing process sequences while using <i>ASTVisitor</i>   |
| Node              | Visited syntax node  |
| Parent node       | Node parent to indicate its parent   |
| Child node        | Node child to indicate its child by node parser  |
| Flag              | Indicating node type<br>(i). Root (node does not have parent and has child or its entry, and it also can indicate starting of program function)<br>(ii). Parent (node has parent and child)<br>(iii). Child (node has parent only, or its leaf node) |
| Syntax classifier | Indicating syntax to distinguish whether its user defined syntax or standard programming language syntax; it can be distinguished by comparing standard programming language syntax data provided and learned in database                            |
| Process           | Indicating process entry (starting process) and exit number of node  |

program preparation, involving program builder process to prepare .bin file for C/C++ and class for java or other languages will be supported in future update. It also processes external source locator for linking process based on the preprocessor statement. Linking and compilation task example for C and C++ program use GNU GCC compiler which produce executable file.exe of linked and final program. Next process is using the executable file to run in DGRADER host machine with input from test case data. The running process of each test case produces an output.txt file. It will be compared with the test

case expected output in database grading criterion. The grading process will be taken after achieving the comparison value of real output and expected output. Finally, final dynamic grading score will be produced.

**4.5. Static Grading.** The inputs of the static grading process include a student program with external sources (if any) and model programs. A student program and its external source will be finalized with program linker-fusion which was already explained before. In this case, both fused student

```

Input:
A: Standardized student program AST
B: Standardized model program AST i
Output: Count of similarity node
Begin:
(1) Step 1: GenerateMap(A, B)
(2)  A → set info tree based on Map key for A tree
(3)  B → set info tree based on Map key for B tree
(4)  global TagMap ← ∅
(5)  global GlobalNodeMap ← ∅
(6)  for each root node  $r \in A \cap B$ 
(7)  do {
      A ← AST of r in A
      B ← AST of r in B
      SimilarityMatching(A, B)
(8) Step 2: SimilarityMatching(A, B)
(9)  ns, cp ← 0
(10) local LocalNodeMap ← ∅
(11) for each (ANode, BNode) ∈ (A, B)
(12)  if (ANode, BNode) = (ATag ANodeNo, BTag BNodeNo)
(13)   TagMap ← TagMap ∪ {ATag ↔ BTag}
(14)   LocalNodeMap ← LocalNodeMap ∪ {ANodeNo ↔ BNodeNo}
(15)  else if (ANode, BNode) = (AChildNo := a op a', BChildNo := b op b')
(16)   SimilarityMatching(a, b)
(17)   SimilarityMatching(a', b')
(18)   if isLocal(a) and is Local (b) and TagMap(a, b) is equal
(19)    LocalNodeMap ← LocalNodeMap ∪ {a ↔ b}
(20)    ns+1, cp+1
(21)   else if
(22)    GlobalNodeMap ← GlobalNodeMap ∪ {a ↔ b}
(23)    cp+1
(24)   else if . . .
(25)   else break
(26) End

```

ALGORITHM 3: AST simple tree matching with tag classifier algorithm.

TABLE 4: Experiment setup for testing dynamic grading.

| Assignment name | Main goal                       | No. of test cases                   | Concurrent process |
|-----------------|---------------------------------|-------------------------------------|--------------------|
| Task 1          | Calculate numbers based on data | Calculate $n$ of $i_n$ data         | 10                 |
| Task 2          | Reverse Fibonacci number        | Print $n$ reversed Fibonacci number | 15                 |

program and model programs will be parsed and produce their program AST. Next process will conduct refactoring of both programs. In our preferred way, model programs are already have been standardized before by the course assignment creator in assignment registration menu to save total time of analysis process. Only student program will be refactored as standardization rule in program matching. This standardization as “one-rule” for avoiding some variances issues will impact the accuracy result in the matching process. After this process, both sides can be compared using program matching using modified AST simple tree matching-pattern algorithm. The task of standardization and program matching will breakdown in Section 4.5.1.

**4.5.1. Program Standardization.** The program standardization algorithm is shown in Algorithm 2. It is an essence work to solve program matching issues such as code variance in programs.

It starts with program AST and continues with traversing and visiting tree process using ASTVisitor. In here, the approach introduces AST data mapper rather than using raw program tree or AST directly to manipulate program which is required in the refactoring process to ease the process of data representation and rebuilding program tree. The AST data mapper stores AST information from ASTVisitor by using map with indexing key (*ASTdata* key written in Table 3). After mapping process, it continues to refactoring process by calling RefactoringCore() to modify or transform the program by using this data map. This procedure will create final *fASTdata*. It will be used for rebuilding final standardized program source code and program tree representation in feedback.

Program refactoring consists of the following steps.

Step 1: removal of problematic code such as syntax error, expression error, and any other with tag problem in *ASTdata* by problem binder which processed while

TABLE 5: Dynamic grading result.

|        | Grading feedback | Result                      |
|--------|------------------|-----------------------------|
|        |                  | Average execution time (ms) |
| Task 1 | Yes              | 20                          |
| Task 2 | Yes              | 136                         |

TABLE 6: Static programming assignment example.

|        | Assignment name          | Main goal  |
|--------|--------------------------|--|
| Task 1 | Calculate numbers (data) | Calculate $n$ of $i_n$ data  |
| Task 2 | Reverse fibonacci number | Print $n$ reversed fibonacci number  |
| Task 3 | Coin changing            | Consider the problem of making change for $n$ cents using the fewest number of coins   |
| Task 4 | <i>Bitonic</i> tour      | Input: $n$ Points: $v_1, \dots, v_n$ on a plane with different $x$ coordinates; $d_{ij}$ : distance between any pair of points ( $v_i, v_j$ ), $i \neq j$ ; output: a <i>bitonic</i> tour with smallest distance |

traversing the program using AST and mapped into *ASTdata* mapper. The problem binder takes advantage of parser to detect syntactical problem. As an example for traversing the C and C++ program using Eclipse CDT *ASTVisitor CPPASTProblem*.

Step 2: identifying and removal of unused function in *ASTdata*. This task deletes its related elements by using start–end index of the node which is already mapped in *ASTdata*.

Step 3: reordering function calling sequences. In this case, main function will become first function call in sequence as the rule of program. The process continues to detect other call function by creating *ASTroot* data mapper. It only targets root type of the node with the root tag filter or classifier. The node is considered as a root node according to flag root detail in Table 3. The result of transformation can be seen from transformation from original AST into root AST.

Step 4: the final step is producing a final standardized program from *ASTroot*.

**4.5.2. Program Matching.** The program matching process requires two program trees as an input which are the student program and model program  $i$ . Both programs are already standardized automatically by the system in the previous process to achieve same structure as one-rule policy in process of program matching.

In program matching, AST simple tree matching is chosen to compare program AST which takes advantage of dynamic programming to calculate maximum node-pair between compared tree program with semantic similarity [29]. However, this algorithm has an issue with overhead memory as it does the code changes when code variance occurred in matching process to make the variance for both compared programs unified. It is also stated in the result test that some code changing activity fail in the process. This will impact the matching similarity accuracy score. Finally, we have improved the approach with AST simple tree matching algorithm, as shown in Algorithm 3.

The algorithm is modified to use advantage of AST tag classifier concept which can be implemented by the method in program AST traversing with any parser. This will solve the overhead and the variance issues.

The improved program matching algorithm has two steps:

Step 1: both programs will be mapped into *ASTMap* which contains key information. It uses root node  $r$  which is already ordered by following the rule in the standardization process. Tag and node are used as matching parameters. *TagMap* is used to take advantage of the AST tag classifier in tag matching comparison. *GlobalNodeMap* and *LocalNodeMap* are used for node element which will be considered equal if encountered in the same tag in the same root node position.

Step 2: after both program trees are mapped into *ASTMap*, the process will invoke *SimilarityMatching()*. Every index data in map will be compared and paired. Accumulate the global *TagMap* and compare *GlobalNodeMap* with *LocalNodeMap* per indexed data in *ASTMap*. Next process is to encounter the exact tag to be compared with it pairs and add into *TagMap* in tag matching task. *LocalNodeMap* used for detection of the identical node in node matching. As stated before, as long as the node follows the same tag and position it will be considered as equal otherwise it will add pattern to find the matching target invoking recursive *SimilarityMatching()* call. For every matching node with same position in local A and B will increase node similarity and pattern count. Total accumulate  $ns$  will be compared to maximum node coverage of both programs. Finally, the final score can be calculated based on how many maximum nodes are covered by  $ns$  in the grading process.

## 5. Experimental Analysis

**5.1. Dynamic Grading.** In dynamic grading, testing conduct with 2 tasks in Table 4. The concurrent process is used to measure system capability to handle concurrent grading process when students finalize their submission at the same time.



TABLE 7: General testing solution program condition.

|            | LOC | Function | Max node | Unused node | Standardization |   |
|------------|-----|----------|----------|-------------|-----------------|---|
|            |     |          |          |             | Transform node  | Program efficiency improvement rate (%) |
| Solution 1 | 12  | 1        | 40       | 6           | 38              | 1.9                                     |
| Solution 2 | 48  | 7        | 94       | 43          | 49              | 52.128                                  |
| Solution 3 | 55  | 9        | 182      | 24          | 158             | 13.19                                   |
| Solution 4 | 120 | 14       | 424      | 35          | 389             | 8.25                                    |

TABLE 8: General testing condition.

|                      | Task 1 | Task 2 | Task 3 | Task 4 |
|----------------------|--------|--------|--------|--------|
| No. of model program | 2      | 2      | 5      | 8      |
| Concurrent process   | 4      | 3      | 7      | 4      |

TABLE 9: General static grading result.

|        | Result           |                    |                 |
|--------|------------------|--------------------|-----------------|
|        | Grading feedback | Analysis time (ms) | Expected result |
| Task 1 | Yes              | 88                 | Yes             |
| Task 2 | Yes              | 90                 | Yes             |
| Task 3 | Yes              | 1022               | Yes             |
| Task 4 | Yes              | 678                | Yes             |

All submission results shown in Table 5 produce grading feedback both successful or error result explanation with the score. The feedback shows success or failure with real and expected output of each test case. In here, testing activity also considers average execution time which measures the time needed by the system to finish the job. Finally, it is successful to achieve dynamic grading mechanism. The concurrent process also can be handled by automatic queuing process and multithreading.

**5.2. Static Grading.** In static grading, programming assignment tasks are shown in Table 6. Task 1 and task 2 are reused in this testing as DGRADER offer flexibility grading assessment method change by updating the assignment configuration. The lecturer just needs to add model programs in the assignment material. Back to testing focus, each sample solution has its behaviour and condition according to Table 7 such as Line of Code (LOC), number of functions, max number of nodes (generated by traversing program AST), and number of unused nodes (code node which are not used in the program).

Standardization data present transforms result from maximum number of nodes to number of transform nodes as automatic transform by program standardization in the system. The number of transform nodes implies the efficiency of the standardized program. The program efficiency improvement rate can be calculated by comparing original maximum number of nodes and number of transform nodes. For example, solution 2 data showed significant 52.128% program efficiency improvement.

Based on Table 8 as the testing condition, the result is shown in Table 9. All solution submitted successfully gives grading feedbacks and results. Expected result means that every variance tolerant achieve full 100% mark as it uses tag

from the AST classifier. It makes every variation occurred in the program solved or tolerated as long as it has true condition from tag and node matching in program matching. The expectation of testing by manual assessment produced. Finally, static grading approach implemented in DGRADER can be found reliable as it is giving good precision of grading score and is also proofed by manual assessment.

**5.3. Reliability and Grading Precision Testing.** This testing is focused in reliability of the final grading score in the static grading assessment. In here, assignment “reverse Fibonacci number” in the previous assignment sample is chosen. The approach uses 1 solution submission and sees its comparison with 4 models (chosen) provided in the assignment. Solution and model programs are shown in Figure 4. The solution program is chosen solution 2 in the previous test, and its condition can be seen in Table 7, and model condition is listed in Table 10.

In here, testing activity hypothesis for model 1 and model 2 will give perfect grading score in the final result. Both the models are nearly same to program solution and also its algorithm which used recursive function to do the job for printing reversed Fibonacci number. The purpose is to check the reliability and precision of logically similar programs by manual assessment. Model 3 is actually .cpp program which uses array approach, and model 4 uses more variables to save values for producing reversed Fibonacci number without recursive function call. They adopt different algorithms to the student solution.

It is clearly evident from Table 11 that similar nodes of solution to models 1 and 2 have 100% final coverage. As stated before, model 1 and model 2 are not exactly the same to the solution. The perfect coverage achieved by variance tolerance successfully tolerates the code variance issues in program matching process. Variance type divided into 3 categories is used based on 9 listed code variance issues in Table 1. In here, the division is based on their related impact into the issue listed variable (name), function (invocation or call, expression, naming), and control structure (compound statement, redundancy, structure order, code format, and algorithm).

Starting analysis with model 1 which is nearly the same, but program solution has code structure data `{int x*, int y*}`. In matching, the pattern increased as it tries to find the node first related to the code with ending in not finding any possible node related to that code. In here, matching still processed and ignored as tolerant in the control structure as the code categorized to impact in the compound statement

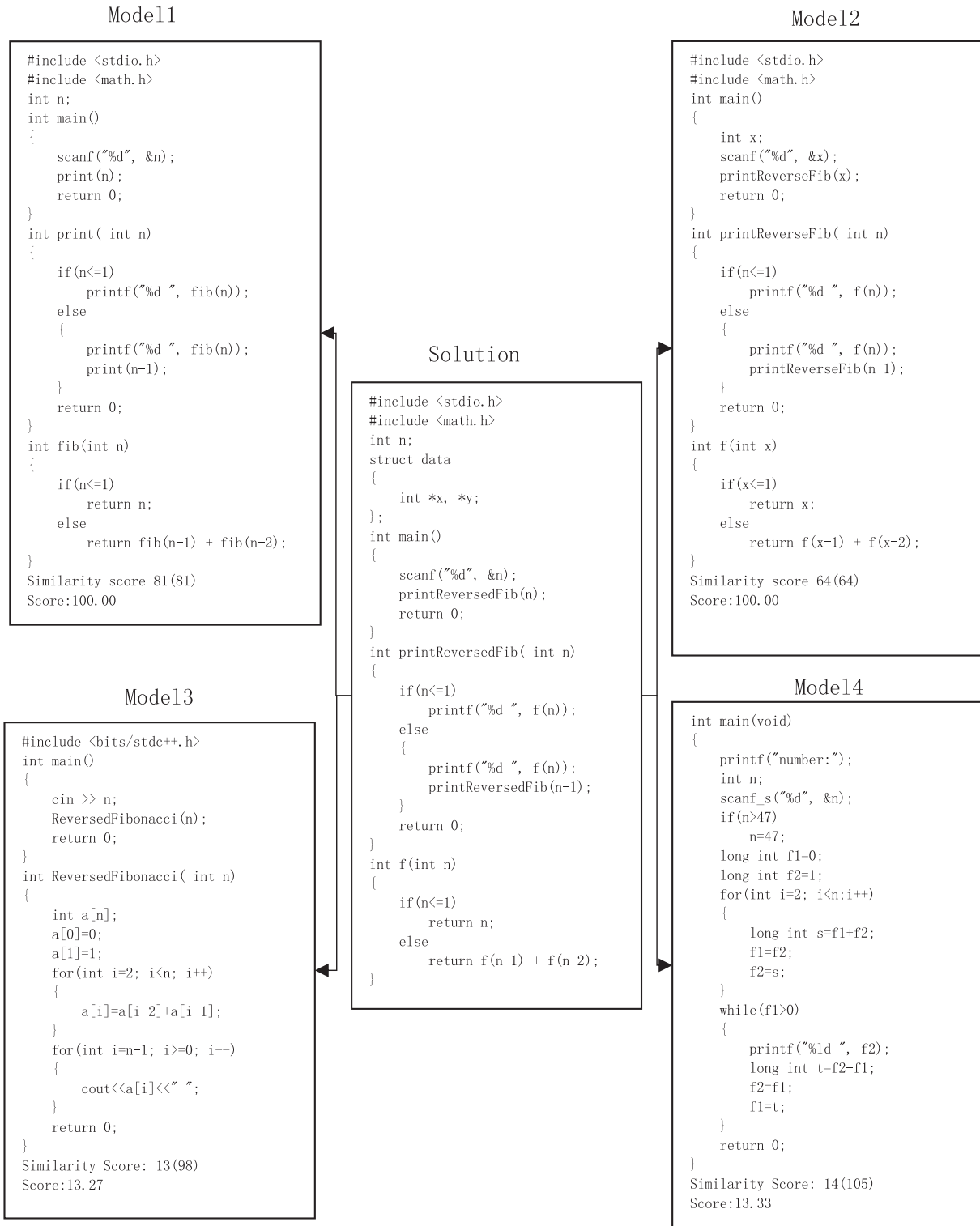


FIGURE 4: Feedback result of program matching trace analysis.

and structure order. It also occurs function naming variance print() and fib() in model compared with printReversedFib() and f() in solution which are successfully tolerated as it is expected to be considered the same. Left remaining codes are exactly the same with solution, which means similarity node coverage of program solution and model 1 is perfect. It

covers 81 nodes of model 1 program with 100% (tolerate all variances) in final coverage.

In model 2, there exists variable variances of node  $x$  which is node  $n$  in program solution. In matching process, it is found that  $x$  corresponds to node  $n$  for the following code. The matching process considered in this node is the same as

TABLE 10: Model program condition.

|         | LOC | Function | Max node | Variance |          |                   |
|---------|-----|----------|----------|----------|----------|-------------------|
|         |     |          |          | Variable | Function | Control structure |
| Model 1 | 48  | 3        | 81       | 2        | 2        | 1                 |
| Model 2 | 48  | 3        | 64       | 1        | 2        | 1                 |
| Model 3 | 25  | 2        | 98       | 1        | 1        | 2                 |
| Model 4 | 22  | 1        | 105      | 4        | 0        | 2                 |

TABLE 11: Static grading precision result.

|                    | Similarity node | Max node | Final coverage (%) | Variable (%) | Variance tolerance |                         | Total pattern matching process |
|--------------------|-----------------|----------|--------------------|--------------|--------------------|-------------------------|--------------------------------|
|                    |                 |          |                    |              | Function (%)       | Control - Structure (%) |                                |
| Model 1            | 81              | 81       | 100                | —            | 100                | 100                     | 394                            |
| Model 2            | 64              | 64       | 100                | 100          | 100                | 100                     | 387                            |
| Model 3            | 13              | 98       | 13, 26             | 100          | 100                | <10                     | 137                            |
| Model 4            | 14              | 105      | 13, 33             | 100          | 100                | <20                     | 159                            |
| Total pattern      |                 |          |                    |              |                    |                         | 1077                           |
| Precision matching |                 |          |                    |              |                    |                         | >98                            |

TABLE 12: Existing similar program comparison (main features).

| Approaches | Supported languages | Complex multfiles analysis | Main features  |  | Grading metric                   |  |
|------------|---------------------|----------------------------|----------------|--|----------------------------------|--|
|            |                     |                            | Platform       | Work mode                                      | Dynamic                          | Static                                 |
|            |                     |                            |                |  |                                  |  |
| Cellidh    | Java, C++           | No                         | Web            | Standalone, competitive learning               | Code correctness                 | Semantic error detection, verification |
| ASSYST     | C/C++               | No                         | Web            | Standalone, competitive learning               | Code correctness, run time check | Code analysis, complexity matrices     |
| Marmoset   | Multilanguage       | No                         | Web            | Standalone                                     | Code correctness                 | Code analysis (model)                  |
| Web-CAT    | Multilanguage       | No                         | Web (Java)     | Standalone, plugins                            | Code correctness, completeness   | Validity check (model)                 |
| eGrader    | Java                | No                         | Desktop (Java) | Standalone                                     | Code correctness                 | Structure matching (model)             |
| AutoLEP    | C/C++               | No                         | Desktop (C#)   | Standalone                                     | Code correctness                 | Similarity matching (model)            |
| Quimera    | C/C++               | No                         | Web            | Standalone, competitive learning               | Code correctness                 | Code quality and accuracy (model)      |
| DGRADER    | C/C++, Java         | Yes                        | Web (Java)     | Standalone, API services, competitive learning | Code correctness, run time check | AST similarity matching (model)        |

expected, as algorithm uses the AST tag classifier in nodes. Besides, structure variance occurs as declaration of  $x$  inside `main()` compared to program solution which is global variable. The matching process is also successful to tolerate this variance and gives similarity 64 node coverages perfectly as expected.

However, models 3 and 4 as expected have low final coverage because both are completely different. The major factor is because both models have distinguished control structure such as algorithm, code format, and related factor which is categorized in this category. Model 3 as stated before is `.cpp` but the system still is capable to compare `.c` program solution with this model which is a plus point. Models 3 and 4 use different approach from solution that uses recursive method. It impacts the matching process

which gives less node coverage and tolerant percentage. The issues can be solved by providing more template program to the system for assignment problem. Finally, final grading score still gives correct score as expected.

*5.4. Case Study Similar Program Comparison Analysis.* The columns in Table 12 refer main features of this objective comparison parameter. From the above parameter, we compare DGRADER as our web-based automatic grading system with the other similar programs which provide flexible grading analysis.

The first key element for comparison analysis is programming language support. Marmoset and Web-CAT are strong in this aspect as they support multilanguages.

TABLE 13: Existing similar program comparison (scoring features).

| Approaches | Categorical grading | Numerical grading | Scoring features        |          |  |     | Ranking      | Plagiarism detection |
|------------|---------------------|-------------------|-------------------------|----------|--|-----|--------------|----------------------|
|            |                     |                   | Error                   | Feedback | Result analysis  |     |              |                      |
| Cellidh    | No                  | No                | Yes                     |          | Yes  | Yes | Yes          |                      |
| ASSYST     | Yes                 | No                | Yes                     |          | Yes  | Yes | Yes          |                      |
| Marmoset   | No                  | No                | No                      |          | Yes  | No  | No           |                      |
| Web-CAT    | No                  | No                | Detailed (highlighting) |          | Code style (highlighting)  | No  | No           |                      |
| eGrader    | No                  | Yes               | Yes                     |          | Program structure  | No  | No           |                      |
| AutoLEP    | No                  | Yes               | Detailed (report)       |          | Semantic analysis  | No  | No           |                      |
| Quimera    | No                  | Yes               | Yes                     |          | Yes  | No  | No           |                      |
| DGRADER    | Yes                 | Yes               | Detailed (report)       |          | Standardization, AST map, transform graph, function call, AST visualization, and matching trace analysis (semantic analysis) | Yes | Configurable |                      |

The second one is complex multifile program analysis feature. Only DGRADER provides complex multifile program analysis for the user to solve one programming problem within more than one modules, libraries, or files. It improves flexibility to solve complex problems with an efficient approach without rewriting program if some libraries exist.

In platform perspective, web platform has become popular consideration rather than becoming a local tool such as AutoLEP and eGrader. Nowadays, everything can be accessed online easily by using browser or mobile device which increases the portability which adds value for flexibility scale.

Next on aspect of work mode, common tools only work as standalone or specific usage. Web-CAT provides plugin for their integration with other platform which require installation. Only DGRADER can work as standalone or integrated with other system by providing its API services which potential for widespread usage. It also delivers with MOOCs with user course management and programming contest platform as feature to support competitive learning. Cellidh, ASSYST, and Quimera also support this competitive learning. It is purposed to increase interest of the user as the content in the system can become user content based.

In grading metrics, compared tools provide flexible grading analysis, which are dynamic and static. However, each approach is different as listed. In dynamic approach, it can be generalized that all approaches are tested with code correctness by using test case input and the output is compared with the expected output. In static, there are various approaches using code analysis. DGRADER uses AST similarity matching with model programs. The approach with model is popularly used by the existing tools. This model is more practical and easier to be measured but requires more models to increase the accuracy and grading precision.

The columns in Table 13 refer the scoring features of this objective comparison parameter. DGRADER comes with various result analyses as their instant feedback and scoring. Instant feedback includes error explanation if any error occurred specially in dynamic grading. In static grading assessment, present feedback such as linked-fused program

source, standardized program source, and mapped into AST data map can be read by the user in result analysis. It also presents transformed graph structure to see change by automatic standardization process, function call sequence, AST visualization of program, and matching trace analysis.

Final scoring provided both in number range 0–100 and categorical score with stars to make the system more interactive. User ranking also provided to stimulate usage and improve user programming and problem solving skill through the reward system with point and level. Plagiarism detection feature is also considered in this scoring feature as it will impact the final grading score. In these compared tools, only Cellidh, ASSYST, and our tool DGRADER provide this feature. The programming assignment creator also can enable plagiarism detection features.

Finally, all comparison analysis can measure to see improvement of each compared tools. Back to DGRADER, our tool currently limited programming language as listed. However, API services will be potential for future development of cross platform integration which means more support of programming language and other features. DGRADER also provides meaningful and informative feedback with novelty complex multifile program analysis feature.

## 6. Conclusions

In this paper, we have presented a novel technique to handle complex multifiles program with flexible static and dynamic grading. It is implemented practically as an automatic grading system platform called DGRADER. In order to deal with multifiles program, the dynamic analysis process takes an advantage of the compiler in linking process to compile complex multifile program. The static analysis process uses the presented complex multifiles program linker-fusion algorithm which parsed preprocessor from program AST to find other external sources. It is successful to link and fuse elements in external sources which are used in the main program. In static analysis, code variance issues in program matching are tackled by the improved algorithm of AST simple tree matching. The AST tag classifier creates tolerant factor of variances from the compared node of solution and

model programs. Program standardization also contributes to transform the program by following the rules in program matching. The results have shown good accuracy in final grading precision as expected in the case of sufficient model programs.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant nos. 61977020 and 61672191), Natural Science Foundation of Heilongjiang Province (Grant no. LH2019F046), Harbin Science and Technology Innovation Talents Research Project (Grant no. 2016RAQXJ013), and Doctoral research fund of Harbin University (Grant no. HUDF2019101).

## References

- [1] S. Li, X. Xiao, B. Basset, T. Xie, and N. Tillman, "Measuring code behavioral similarity for programming and software engineering education," in *Proceedings of the ACM 38th IEEE International Conference on Software Engineering Companion*, pp. 501–510, Austin, TX, USA, May 2016.
- [2] S. Gulwani, I. Radiček, and F. Zuleger, "Automated clustering and program repair for introductory programming assignments," in *Proceedings of the 39th ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI 2018)*, pp. 465–480, New York, NY, USA, June 2018.
- [3] T. Wang, J. Xu, X. Su, C. Li, and Y. Chi, "Automatic debugging of operator errors based on efficient mutation analysis," *Multimedia Tools and Applications*, vol. 78, no. 21, pp. 29881–29898, 2019.
- [4] D. G. Kay, T. Scoot, P. Isaacson, and K. A. Reek, "Automated grading assistance for student program," *ACM SIGCSE Bulletin*, vol. 26, no. 1, pp. 381–382, 1994.
- [5] P. Li and L. Toderick, "An automatic grading and feedback system for e-learning in information technology education," in *Proceedings of the ASSE Annual Conference and Exposition for Emerging Computing and Information Technologies*, pp. 1–11, Seattle, Washington, June 2015.
- [6] C. Wilcox, "Testing strategies for the automated grading of student program," in *Proceedings of the 47th ACM Technical Symposium on Computing Science Education—SIGCSE '16*, pp. 437–442, Memphis, TN, USA, March 2016.
- [7] J. Qi, G. Jiang, G. Li, Y. Sun, and B. Tao, "Intelligent human-computer interaction based on surface EMG gesture recognition," *IEEE Access*, vol. 7, pp. 61378–61387, 2019.
- [8] G. Li, L. Zhang, Y. Sun, and J. Kong, "Towards the sEMG hand: internet of things sensors and haptic feedback application," *Multimedia Tools and Applications*, vol. 78, no. 21, pp. 29765–29782, 2019.
- [9] G. Conole and B. Warburton, "A review of computer-assisted assessment," *ALT-J*, vol. 13, no. 1, pp. 17–31, 2005.
- [10] K. M. Ala-Mutka, "A Survey of automated assessment approaches for programming assignments," *Computer Science Education*, vol. 15, no. 2, pp. 83–102, 2005.
- [11] P. Ilhantola, T. Ahoniemi, V. Karavirta, and O. Seppala, "Review of recent systems for automatic assessment of programming assessment," in *Proceedings of the 10th Koli Calling International Conference on Computing Education Research—Koli Calling '10*, pp. 86–93, Koli, Finland, October 2010.
- [12] T. Wang, X. Su, and P. Ma, "Program normalization for removing code variations," in *Proceedings of the 2008 IEEE International Conference on Computer Science and Software Engineering*, pp. 306–309, Hubei, China, December 2008.
- [13] T. Wang, X. Su, and P. Ma, "Function inlining algorithm for program analysis," in *Proceedings of the 2009 IEEE International Conference on Computational Intelligence and Software Engineering*, pp. 1–4, Wuhan, China, December 2009.
- [14] S. M. Arifi, A. Zahi, and R. Benabbou, "Semantic similarity based evaluation for C programs through the use of symbolic execution," in *Proceedings of the 2016 IEEE Global Engineering Education Conference*, pp. 826–833, Abu Dhabi, UAE, April 2016.
- [15] A. N. Jacobvitz, A. D. Hilton, and D. J. Sorin, "Multi-program benchmark definition," in *Proceedings of the 2015 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS)*, pp. 72–82, Philadelphia, PA, USA, March 2015.
- [16] K. Matthews, T. Janicki, L. He, and L. Patterson, "Implementation of an automatic grading system with an adaptive learning component to affect student feedback and response time," *Journal of Information System Education*, vol. 23, no. 1, pp. 71–83, 2012.
- [17] M. Pozenel, L. Furst, and V. Mahnic, "Introduction of the automated assessment of homework assignments in a university-level programming course," in *Proceedings of the 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 761–766, Opatija, Croatia, May 2015.
- [18] T. Wang, X. H. Su, and P. J. Ma, "Semantic similarity-based grading of student programs," *Information Software Technology*, vol. 49, no. 2, pp. 17–31, 2007.
- [19] D. Fonte, D. Cruz, A. L. Gancarski, and P. R. Henriques, "A flexible dynamic system for automatic grading of programming exercises," in *Proceedings of the 2nd Symposium on Language, Applications and Technologies*, pp. 129–144, Dagstuhl, Germany, 2013.
- [20] S. D. Benford, E. K. Burke, E. Foxley, and C. A. Higgins, "The Ceilidh system for the automatic grading of students on programming courses," in *Proceedings of the 33rd Annual on Southeast Regional Conference ACM-SE 33*, Clemson, SC, USA, March 1995.
- [21] D. Jackson and M. Usher, "Grading student programs using ASSYST," in *Proceedings of the 28th SIGCSE Technical Symposium on Computer Science Education*, pp. 335–339, San Jose, CA, USA, February 1997.
- [22] J. Spacco, D. Hovemeyer, W. Pugh, F. Emad, J. K. Hollingsworth, and N. Padua-Perez, "Experiences with marmoset," *ACM SIGCSE Bulletin*, vol. 38, no. 3, pp. 13–17, 2006.
- [23] S. H. Edwards and M. A. Perez-Quinones, "Web-CAT," *ACM SIGCSE Bulletin*, vol. 40, no. 3, p. 328, 2008.
- [24] F. Alshamsi and A. Elnagar, "An automated assessment and reporting tool for introductory Java programs," in *Proceedings of the 2011 International Conference on Innovations in*

- Information Technology (IIT)*, pp. 324–329, Abu Dhabi, UAE, April 2011.
- [25] T. Wang, X. Su, P. Ma, Y. Wang, and K. Wang, “Ability-training-oriented automated assessment in introductory programming course,” *Computers & Education*, vol. 56, no. 1, pp. 220–226, 2011.
- [26] D. Fonte, I. V. Boas, D. Cruz, A. L. Gancarski, and P. R. Henriques, “Program analysis and evaluation using quimera,” in *Proceedings of ICEIS*, pp. 209–219, Wroclaw, Poland, June 2012.
- [27] T.-H. Wang, “Web-based dynamic assessment: taking assessment as teaching and learning strategy for improving students’ e-Learning effectiveness,” *Computers & Education*, vol. 54, no. 4, pp. 1157–1166, 2010.
- [28] D. Muñoz de la Peña, F. Gómez-Estern, and S. Dormido, “A new internet tool for automatic evaluation in control systems and programming,” *Computers & Education*, vol. 59, no. 2, pp. 535–550, 2012.
- [29] I. Neamtii, J. S. Foster, and M. Hicks, “Understanding source code evolution using abstract syntax tree matching,” in *Proceedings of the 2005 International workshop on Mining Software Repositories*, pp. 1–5, Saint Louis, MO, USA, May 2005.

## Research Article

# A Hemodynamic-Based Evaluation of Applying Different Types of Coronary Artery Bypass Grafts to Coronary Artery Aneurysms

Haoran Wang,<sup>1,2</sup> Hitomi Anzai,<sup>1</sup> Youjun Liu,<sup>3</sup> Aike Qiao ,<sup>3</sup> Jinsheng Xie,<sup>4</sup> and Makoto Ohta <sup>1,2,5</sup>

<sup>1</sup>Institute of Fluid Science, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai, Miyagi 980-8577, Japan

<sup>2</sup>Graduate School of Biomedical Engineering, Tohoku University, 6-6 Aramaki-aza-aoba, Aoba-ku, Sendai, Miyagi 980-8579, Japan

<sup>3</sup>College of Life Science and Bioengineering, Beijing University of Technology, No. 100, Pingleyuan, Chaoyang District, Beijing 100022, China

<sup>4</sup>Department of Cardiac Surgery, Beijing Anzhen Hospital, Capital Medical University, Beijing Institute of Heart Lung and Blood Vessel Diseases, Beijing 100029, China

<sup>5</sup>ELyTMax UMI 3757, CNRS-Université de Lyon, Tohoku University, Sendai, Miyagi 980-8579, Japan

Correspondence should be addressed to Makoto Ohta; ohta@biofluid.ifs.tohoku.ac.jp

Received 28 March 2020; Revised 6 May 2020; Accepted 11 May 2020; Published 10 June 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Haoran Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Coronary artery bypass grafts (CABGs), including saphenous vein grafts (SVGs) or left internal mammary artery (LIMA) grafts, are recently applied to treat coronary artery aneurysm (CAA). Surgical outcomes are considered to be related to surgical strategies (types of the bypass graft and whether CAA ligated or not) and the size of the CAA (usually characterized by diameter). However, the understanding of the relationship between the surgical outcomes and the abovementioned factors is limited. Previous studies related to CABG treatments have shown hemodynamic studies could help evaluate surgical outcomes through graft mass flow rate, wall shear stress (WSS), and oscillatory shear index (OSI). It is believed that the hemodynamic study of applying CABGs to CAA, which is not studied yet, could help us understand the different CABG surgeries. The aim of the study was to evaluate the hemodynamic differences among different surgical methods. To do this, eight three-dimensional models were constructed, representing application of SVGs and LIMA grafts to CAAs (whether ligated or not) with diameters two, three, and five times the normal diameter, to perform computational fluid dynamics (CFD) simulation. The lumped-parameter model (LPM) was coupled to the boundary of the 3D models which increase the complexity of the simulation, but it can ensure the stability of the simulation boundary conditions. The results show that SVG (no matter whether ligated or not) hemodynamic characteristics are positive, with an average high graft mass flow rate of 70 ml/min, an average WSS of 0.479 Pa, and a low OSI of 0.001. LIMA with CAA ligation has the same characteristics with higher WSS (average 1.701 Pa). The hemodynamic characteristics of LIMA without CAA ligation are negative, including high reverse mass flow rate and high OSI (0.367). The results indicate that the surgical outcomes of LIMA with CAA ligation are likely to be the best among these models. The surgical outcomes of LIMA without CAA ligation seem to be undesirable due to the high reverse mass flow and high OSI. The CAA diameter may not have a significant effect on surgical outcomes.

## 1. Introduction

Coronary artery aneurysms (CAAs) are cardiovascular disease, which is defined as dilatation of the coronary artery exceeding more than 50% of the reference vessel diameter, and the incidence rate is approximately 0.3–5.3% [1, 2]. It

was learned that up to one-third of CAAs are associated with obstructive coronary artery disease and have been associated with myocardial infarction, arrhythmias, or sudden cardiac death [3]. Recently, coronary artery bypass graft (CABG) surgery is usually selected to treat myocardial ischemia and complications caused by CAAs [4–6]. However, because of

the rarity of CAA and the variety of surgeries, our understanding of applying CABGs to CAA is limited, most of which comes from case reports [7–11]. With the increase in patients and more and more applications of surgery, it is necessary to conduct a research on applying CABGs to CAA to help us understand the surgery.

There are usually two different surgical options and an important parameter to be considered when implementing CABG applying to CAA. The different surgical options include the types of CABGs and whether the CAA is ligated or not. Saphenous vein grafts (SVG) and left internal mammary artery (LIMA) grafts, harvested from the patients' own body, are the main choice for the bypass graft of CABG surgery [12]. SVG and LIMA have their own advantages and disadvantages, and there is no uniform standard. However, the SVG is the most widely used. Compared with LIMA, the SVG has the advantage of ease of use, sufficient length, and accessibility. Especially, for patients with a diffusely atherosclerotic ascending aorta, SVG composite Y-grafts may be necessary, particularly when there is some contraindication to the use of arterial grafts and severe (>70%) stenosis [13–15]. The CAA ligated or nonligated can cause different degrees of damage to blood vessels. Ligation is to cut off the original coronary arteries, which will cause huge structural damage to the cardiovascular system. Without ligation, it will only cause anastomotic damage and no other structural damage. The important parameter to be considered is the size of the CAA, which is usually characterized by diameter. Changes in the size of the CAA diameter could affect the complications, which means the surgical strategy may be affected.

To the best of our knowledge, because of scarcity of CAA and complexity of factors affecting the surgical outcomes of surgery, there are no research studies conducted on applying CABGs to CAA to investigate the differences in surgical outcomes caused by different types of bypass grafts, CAA ligated or not. The effect of CAA diameter on different surgical methods is also unknown. The study of hemodynamics provides an effective method for us to solve the abovementioned problems. Previous studies have pointed out that hemodynamic study of applying CABGs to treat other diseases such as coronary stenosis by using computational fluid dynamics (CFD), which has attracted wide attention in recent years and has many applications in the vascular field [16–18], has made progress. Also, those research studies confirm that surgical outcomes correlate with some important hemodynamic characteristics including mass flow rate, wall shear stress (WSS), and oscillatory shear index (OSI) [19–24]. Therefore, it is reasonable to believe that hemodynamic study of applying CABGs to CAA can be used to evaluate the surgical outcomes.

The aim of this study was to evaluate the hemodynamic differences between two different surgical methods (types of CABGs, whether CAA ligated or not) and the effect of CAA diameter on the hemodynamics of different surgical methods through the CFD method. It could allow us to evaluate different surgical procedures of CAA and make preliminary predictions and deduction for different surgical plans based on these hemodynamic characteristics from a

hemodynamic perspective. This research may be the first hemodynamic study that evaluates the hemodynamic characteristics of different types of bypass grafts on CAA of different diameters and makes a rank of treatment procedure based on the hemodynamic characteristics. This study may provide information to clinicians for better predicting of surgery and making more reliable treatment decisions.

## 2. Materials and Methods

**2.1. Model Development.** A patient-specific 3D geometric anatomical model was reconstructed based on medical images that were obtained by computed tomography angiography (CTA, Siemens). The CTA images of the patient, whose information was anonymized in this research, were provided by Anzhen Hospital with  $512 \times 512$  pixels of each image, 450 slice images, and 1 mm gap between adjacent slice. This study has been approved by the Medical Ethics Committee of Beijing Anzhen Hospital and Tohoku University. The construction processing of the 3D model was done by using Mimics (Materialise NV, BE) through both manual and threshold segmentation. Three different diameters of CAA, including the diameters of two, three, and five times compared with the adjacent coronary artery, were applied to the left anterior descending artery by using Geomagic Freeform (3D system, US) software. Virtual bypass surgery was implemented on these models including SVG and LIMA bypass grafts. Two more virtual bypass surgery models (SVG and LIMA) in which CAA was ligated were also established. Figure 1 shows the details of all different 3D models which were named model 1–8, respectively.

**2.2. Computational Model and Properties of Fluid.** The finite element analysis method will be adopted in which all models need to be preprocessed, including preprocessing of the inlets and outlets and the smooth processing of the area of interest. Preprocessing models need to be divided into meshes, which were hexahedral meshes that were controlled by means of size control with the software ANSYS Meshing (ANSYS, US), to generate the fluid computational models. The mesh of the area of interest is refined in order to get more accurate results in these areas. The mesh independency was analyzed to guarantee the reliability of the simulation results, which needed the number of elements and nodes to be large enough [25]. Table 1 shows the results of the mesh independency test. The differences in WSS between the chosen mesh were less than 0.5%. The calculation result can be considered stable. Table 2 shows the details of the number of elements and nodes of all 8 models in this paper. Blood flow simulations were performed with the ANSYS-CFX (ANSYS, US) Navier–Stokes solver. In the simulation, we assumed that the wall was rigid and stationary. This study assumed that blood was an incompressible Newtonian fluid with a viscosity of  $0.0035 \text{ Pa}\cdot\text{s}$  and a density of  $1050 \text{ kg/m}^3$ . Simulations were run for 3 cardiac cycles. We selected  $0.0025 \text{ s}$  time step to satisfy the calculation of stability conditions. The maximum convergence RMS residual of simulation was set to 0.0001.



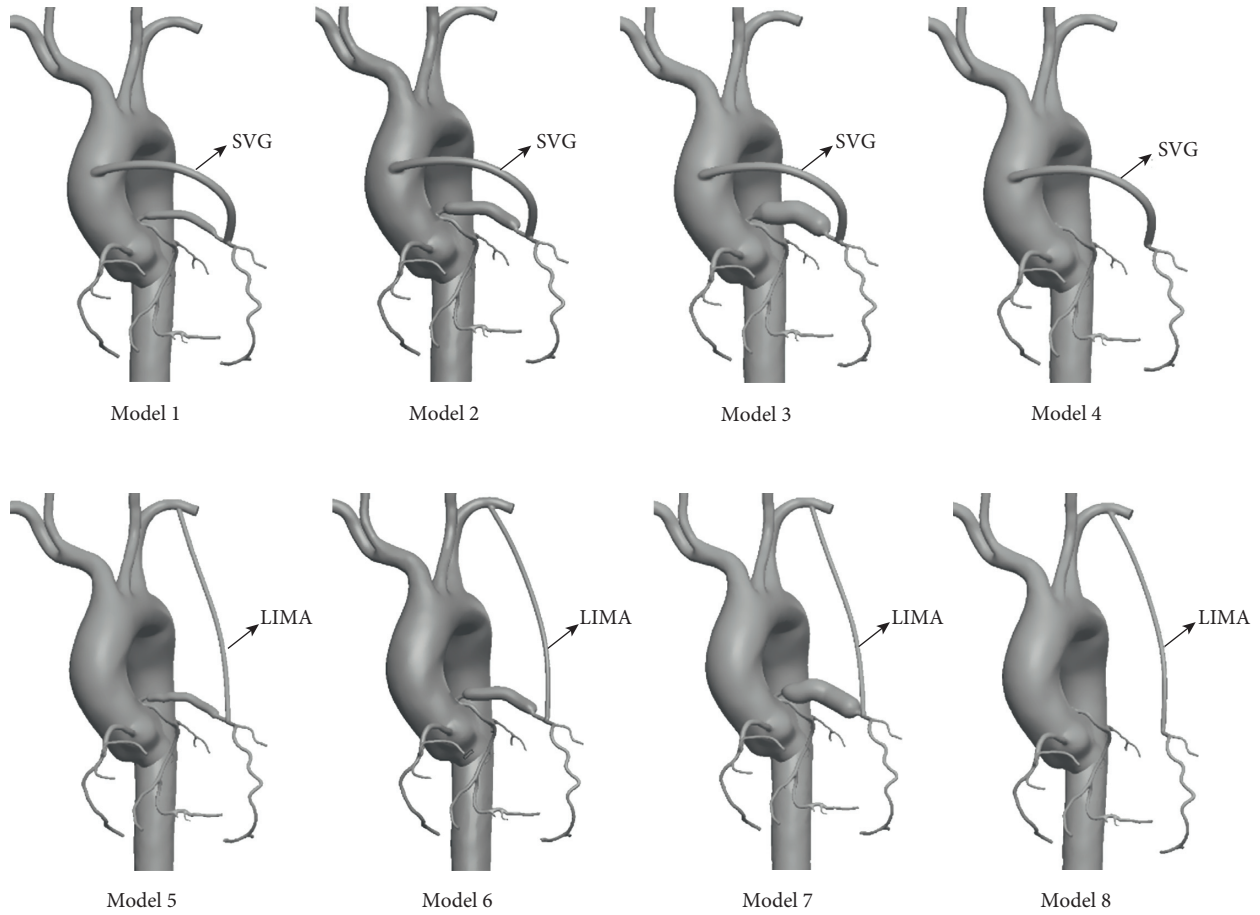


FIGURE 1: All eight 3D CABG surgery models which describe the location and the size of the CAA and the way of how to perform the bypass surgery. Models 1–4: bypass graft is SVG and the diameter of CAA is 2x, 3x, 5x, and 0 (CAA ligated), separately. Models 5–8: bypass graft is LIMA and the diameter of CAA is 2x, 3x, 5x, and 0 (CAA ligated), separately.

TABLE 1: The results of the mesh independency test on WSS and pressure.

| Number of meshes | Percent change of the parameter values | SVG            |                  | LIMA           |                 |
|------------------|--|----------------|------------------|----------------|-----------------|
|                  |  | WSS (Pa)       | Pressure (mmHg)  | WSS (Pa)       | Pressure (Pa)   |
| 200,000          |  | 0.626          | 77.912           | 0.639          | 79.038          |
| 400,000          | $\epsilon$                             | 0.599<br>4.30% | 86.293<br>10.70% | 0.623<br>2.50% | 86.395<br>9.30% |
| 600,000          | $\epsilon$                             | 0.621<br>3.60% | 82.629<br>4.20%  | 0.62<br>0.50%  | 82.547<br>4.50% |
| 800,000          | $\epsilon$                             | 0.61<br>1.80%  | 84.324<br>2.00%  | 0.608<br>1.90% | 84.672<br>2.60% |
| 1,000,000        | $\epsilon$                             | 0.607<br>0.50% | 84.197<br>0.20%  | 0.609<br>0.16% | 84.925<br>0.30% |

$\epsilon$  is the percentage change of the parameter value (WSS and pressure) due to the increase in the number of meshes.

**2.3. Boundary Conditions: Lumped-Parameter Model Boundary Conditions.** Boundary conditions of computational simulation play an important role that can affect the accuracy of the results and determine whether the calculation is successful. The ideal situation is to obtain the real human beings' physiological parameters as the boundary conditions of the simulation calculation. However, obtaining coronary artery physiological parameters is not easy due

to its small size, numerous branches, and distribution on the heart's surface. Therefore, in order to obtain life-like boundary conditions, we selected the lumped-parameter model (LPM), which is a circuit model that can simulate the blood flow and pressure through current and voltage, to provide boundary conditions for the CFD simulation [25–27]. One reason for choosing LPM is its simplicity. Complex coronary branches can be simulated by simple

TABLE 2: The values of nodes and elements of all models.

|          | Model 1   | Model 2   | Model 3   | Model 4   | Model 5   | Model 6   | Model 7   | Model 8   |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Nodes    | 1,007,389 | 1,016,509 | 1,145,628 | 1,003,758 | 1,004,659 | 1,018,965 | 1,159,365 | 1,007,932 |
| Elements | 1,238,694 | 1,335,889 | 1,369,378 | 1,218,463 | 1,239,574 | 1,339,365 | 1,359,436 | 1,228,346 |

circuit modules while retaining the characteristics of coronary blood supply. On the other hand, less demand for computing resources, which means better adaptability and lower consumption, is meaningful.

LPM consists of three modules, which are arterial module, coronary module, and heart module. The arterial module consists of many branches which base a three-element model (Windkessel RCR model) [28] structure that consists of two resistors and one capacitor. The coronary module consists of the arterial module and coronary microcirculation module. The main feature of coronary microcirculation is the resistance to drastic changes caused by the contraction of the heart. The myocardial compliance ( $C_{i-micro}$ ) and microcirculation resistance ( $R_{v-micro}$ ) were used to simulate this feature, where the cathode of the capacitor is loaded with intramyocardial pressure to simulate the pressure generated by the contraction of the heart during the systolic phase. In the LPM heart model, all the simulation equations of the left ventricle is based on the previous research [25] including  $C(t)$ ,  $E(t)$ ,  $E_{max}$ ,  $E_{min}$ ,  $E_n(t_n)$ , and  $t_n$ , and we determined the specific value of each component including resistors and capacitors, which are adjusted to match the patient's blood pressure.

Based on the abovementioned 0D model and 3D models, the multiscale 0D/3D coupling model used in this research was constructed. Figure 2 shows coupling interface details of the 0D/3D coupling model. All 3D models were coupled with the same 0D model whose parameter values adopt the values of the normalized coronary model, which is confirmed by observing that the geometric structure of the distal coronary artery was not changed. Models sharing the same LPM can also exclude other factors to ensure that the change of the values of hemodynamic parameters is only related to the geometric structure.

**2.4. Hemodynamic Parameters.** Hemodynamics can reveal the characteristics of the blood flow. We will focus on several key hemodynamic parameters including mass flow rate, WSS, and OSI. The WSS represents the force per unit area of fluid applied along the wall surface in the direction of the local tangent plane [29]. Formula (1) is as follows:

$$WSS = \tau_\omega = \mu \left( \frac{\partial u}{\partial y} \right)_{y=0}, \quad (1)$$

where  $\mu$  is the dynamic viscosity,  $u$  is the velocity parallel to the wall, and  $y$  is the distance to the wall. OSI quantifies WSS direction changes. OSI is calculated using the following formula:

$$OSI = \frac{1}{2} \left( 1 - \frac{\left| \int_0^T \vec{\tau}_\omega dt \right|}{\int_0^T |\vec{\tau}_\omega| dt} \right). \quad (2)$$

### 3. Results

In this finite element simulation, we evaluated some key hemodynamic parameters which could help evaluate surgery, including mass flow rate, WSS, and OSI on the bypass graft.

**3.1. Mass Flow and Pressure Drop.** The values of mass flow through the bypass grafts and the pressure difference between the upstream and downstream of the bypass grafts over a cardiac cycle are shown in Figure 3. The results show that the mass flow rate of the SVG (no matter) is much higher than that of LIMA, in which the peak value is up to 120 ml/min and the mean value is 50 ml/min, except the LIMA model with ligation of CAA. The mass flow rate of the LIMA in model 8 with ligation of CAA is almost the same as that of the SVGs. The mean mass flow rate of the LIMA in models 5–7 without ligation of CAA is almost 0. The reverse flow which means the blood flows from the bottom to the top of the LIMA can be found in these models without ligation of CAA during the systolic phase of the heart. The negative pressure difference is found in the LIMA models without CAA ligation (models 5–7), while the pressure difference of other models is always positive. Pressure difference is the main driving force of liquid in fluid simulation, which may indicate the main cause of the reverse flow.

**3.2. Wall Shear Stress.** WSS is a force applied to a tangent plane which indicates the per unit area force exerted by the fluid [30]. The values of WSS on SVGs and LIMAs of all models over a cycle and WSS distribution on the SVGs and LIMAs of all models (peak point) are shown in Figures 4 and 5. The mean area WSS on the bypass grafts of models 1–8 over a cycle is 0.477, 0.483, 0.516, 0.443, 0.851, 0.877, 0.896, and 1.701 Pa, respectively. Under the same conditions, WSS on the LIMA is higher than that of the SVG. WSS on the LIMA with CAA ligation (model 8) is especially higher.

**3.3. Oscillatory Shear Index.** OSI is a parameter that can quantify the change in direction and magnitude of the WSS [31]. The distribution of OSI on all models is shown in Figure 6. The mean area values of OSI on SVGs and LIMAs of all models are 0.001, 0.001, 0.001, 0.001, 0.343, 0.281, 0.476, and 0.001, respectively. The values of OSI on SVGs (no matter whether CAA is ligated or not, models 1–4) and the values of OSI on LIMA with CAA ligation (model 8) are almost 0 which means the direction of blood flow has almost no change. In contrast, the values of OSI on LIMAs without CAA ligation (models 5–7) are decades of orders of magnitude higher than the same condition of SVGs (no matter whether CAA is ligated or not).

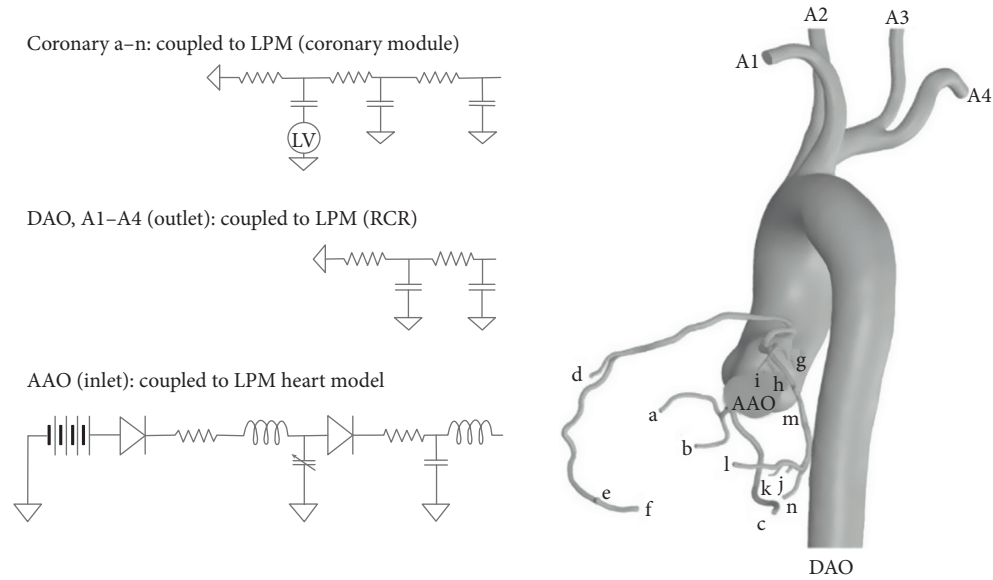


FIGURE 2: The construction of interfaces of the 0D/3D coupling model. In the coupled system, the coronary module of LPM is coupled with the coronary arteries of the 3D model, the LPM heart module is coupled with the aortic inlet of the 3D model, and the common artery LPM module (RCR) is coupled with other branch arteries of the 3D model.

## 4. Discussion

CABG (including SVG and LIMA) surgical operation nowadays is generally selected to treat CAA. Moreover, whether CAA is ligated or not is another choice. Surgical outcomes differ depending on the types of bypass grafts and diameter of CAA (ligated or not). Hemodynamic characteristics (mass flow rate, WSS, and OSI) of different types of CABG surgeries (SVG and LIMA) applied to CAA (different diameters, ligated or not), which were related to surgical outcomes based on the previous studies, are obtained through the CFD method.

We found that the hemodynamics of CAA ligation models is better than those of models without CAA ligation. When CAA was not ligated, the SVG has better hemodynamic characteristics than LIMA. When the same type of bypass graft is used, the change of CAA diameter has basically no effect on the hemodynamic characteristics of the bypass graft.

We deduced that surgical outcomes of LIMA with CAA ligation may be the best and surgical outcomes of the SVG may be superior to LIMA without CAA ligation. Surgical methods, including the type of bypass graft and CAA ligated or not, may be the main factors affecting the outcomes of the operation, and CAA diameter could be a minor factor.

**4.1. Characteristics of Mass Flow Rate of Bypass Grafts.** Mass flow rate, which is one of the few parameters that can be measured directly during the procedure, is one of the key hemodynamic parameters for clinical surgery. Most of the time, this indicator is the most direct indicator for judging the success of surgery [19, 32]. There are no significant differences in the mass flow rate of SVG between different diameters of CAA. The mass flow curves of SVG with and

without CAA ligation are similar in which the minimum and maximum values are 50 ml/min and 120 ml/min. The mass flow curves of LIMA models 5–7, without CAA ligation, are similar in which the minimum and maximum values are –50 ml/min and 60 ml/min. There is significant reverse flow during the systolic phase. Studies have shown that reverse flow is harmful to bypass grafts and is associated with the string phenomenon which indicates the failure of the bypass grafts [19, 20]. A reverse flow value of  $\geq 3.0\%$  can be considered as a cutoff value that predicts early graft failure [20]. The reverse flow value of 43.5% without CAA ligation is much higher than 3.0% in this research. Therefore, early graft failure may happen in LIMA without CAA ligation. However, by comparison, mass flow rate of LIMA with CAA ligation is much higher than that of LIMA without CAA ligation, in which the minimum and maximum values are 30 ml/min and 100 ml/min and no reverse flow exists. Mass flow rates of SVGs with and without CAA ligation and LIMA with CAA ligation are usually the expected results which are with higher mass flow and no reverse flow.

The inlet locations and the diameters of the grafts are the major geometrical differences between SVG and LIMA grafts in this research. On the one hand, the inlet of an SVG is closer to the aortic root than that of an LIMA graft. Therefore, the pressure delay of SVGs is less than that of LIMAs. On the other hand, the location close to the aortic root ensures that the pressure upstream of the bypass graft is close to the aortic pressure which can keep the pressure difference between the upstream and downstream of the bypass graft positive. The inlet of the LIMA is far from the root of the aorta which caused a negative pressure difference across the LIMA. The pressure difference has been shown in Figure 3 which can confirm our point of view. Except for LIMA without CAA ligation, the pressure difference of the bypass grafts in other models is always positive in the cardiac

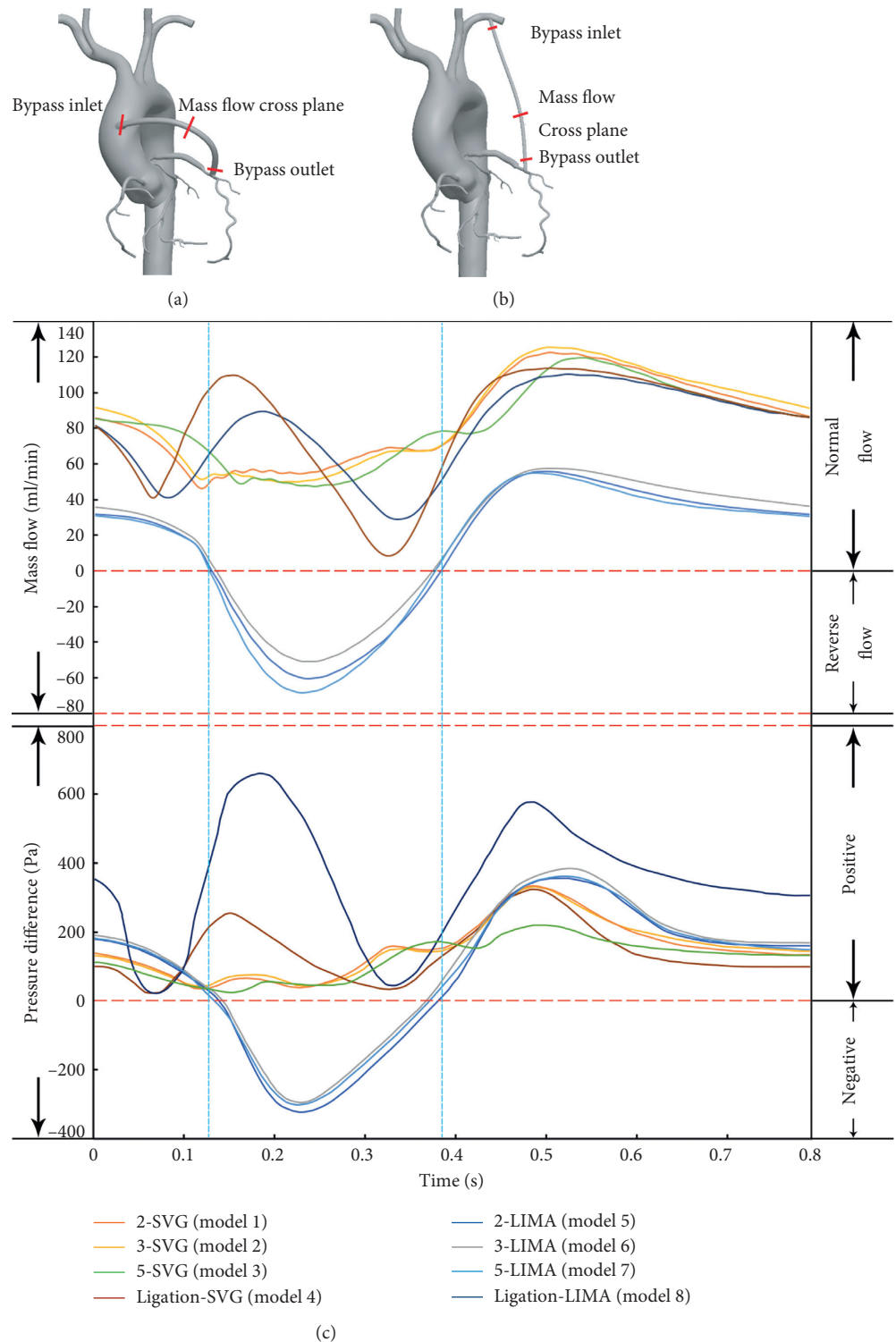


FIGURE 3: The mass flow rate depending on times of the bypass grafts through the cross plane and the pressure difference between the bypass inlet and the bypass outlet. (a) Representation of the SVG virtual surgery model and three locations on the SVG. (b) Representation of the LIMA virtual surgery model and three locations on the LIMA. (c) The mass flow rate through the cross plane of the bypass grafts in all models, as shown in (a) and (b), and the pressure difference between the bypass inlet and bypass outlet in all models, as shown in (a) and (b).

cycle. In the models of LIMA without CAA ligation, the significant negative pressure difference was found which might be the major reason for the reverse flow during the

systole period in an LIMA graft. A larger diameter means less resistance. This may be the main reason why the mass flow of SVG is higher than that of LIMA with CAA ligation.

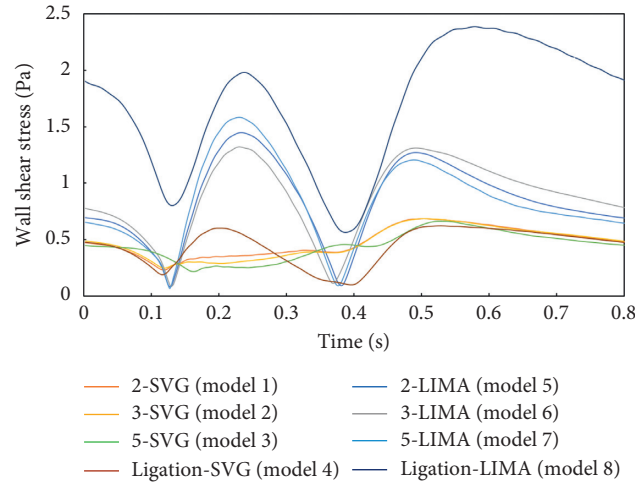


FIGURE 4: The WSS depending on times of the bypass grafts (SVG and LIMA) in a cardiac cycle.

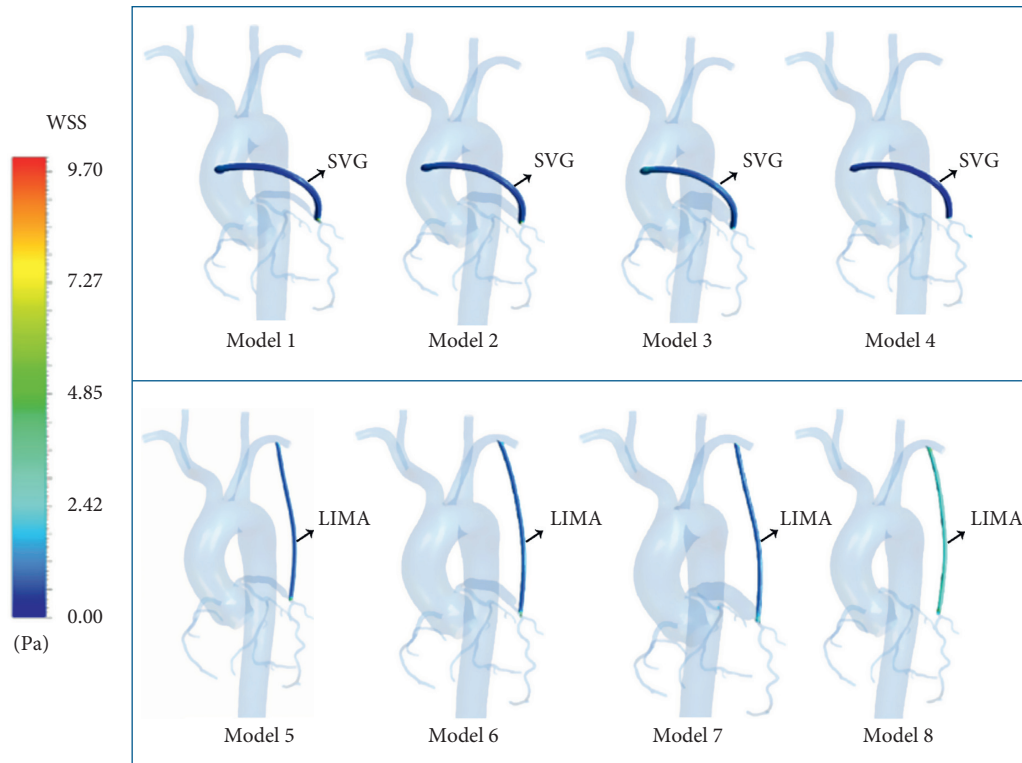


FIGURE 5: The WSS distribution on the bypass grafts (SVG and LIMA) at the peak point ( $t=0.24$  s) in a cardiac cycle.

**4.2. Characteristics of WSS of Bypass Grafts.** WSS is related to some important physiological processes of endothelial cells, such as the formation of important compounds [21, 22]. According to previous research including simulation and experimental research, WSS values have different effects on physiological processes of endothelial cells when they are in different ranges. According to the experimental reports, the results from experiments in parallel-plate flow chambers revealed that a low WSS value of 0.4 Pa induced extensive platelet aggregation [33]. This

abnormal behavior is considered to have the potential to cause vascular diseases. The simulation study also reached a similar conclusion, that is, when the WSS is less than 0.4 Pa, the blood vessels are prone to lesions [34, 35]. Simulation studies also found that the WSS interval which maintains normal physiological functions of blood vessels is generally considered to be 1–7 Pa. WSS values ranging from 0.4–1 Pa is usually considered a transition zone where the trend of blood vessel development cannot be easily predicted. Therefore, even if the WSS is greater than 0.4, we cannot

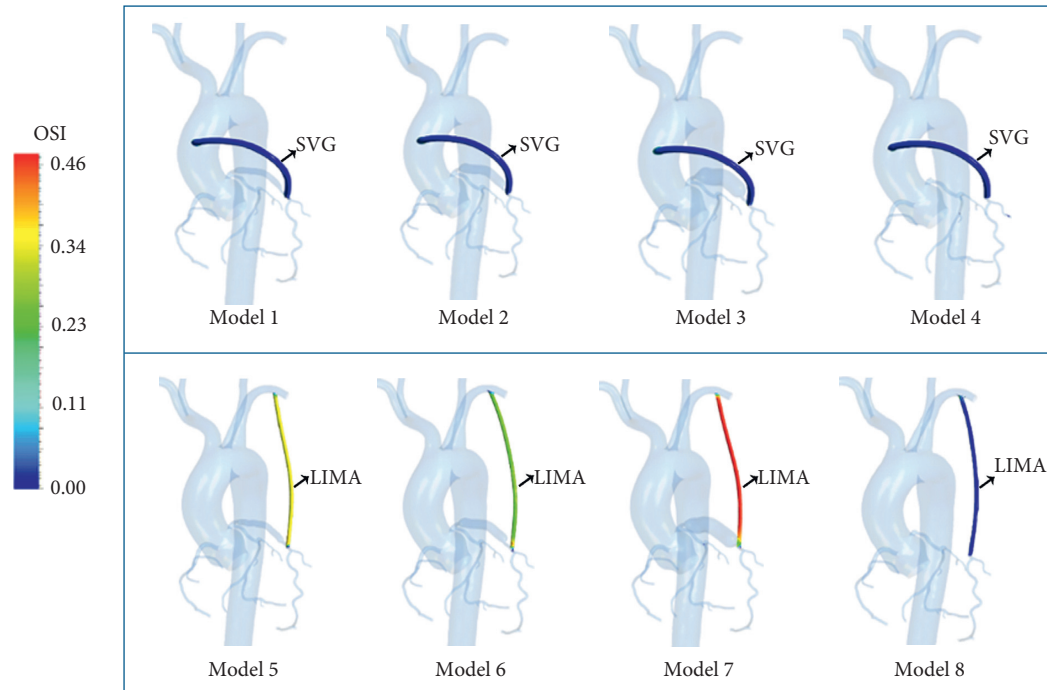


FIGURE 6: The OSI distribution on the bypass grafts (SVG and LIMA) in a cardiac cycle.

guarantee that the function of the blood vessels will not be affected at all. Our research results are as follows. WSS in SVGs (with and without CAA ligation, models 1–4) is similar, with an average value of 0.479 Pa. WSS in LIMAs (without CAA ligation, models 5–7) is similar, with an average value of 0.874 Pa. WSS in LIMA (with CAA ligation, model 8) is much higher, with a value of 1.701 Pa. In the simulation results, only the WSS value in LIMA with CAA ligation is in the range of 1–7 Pa and all others are in the range of 0.4–1 Pa. Therefore, LIMA with CAA ligation may be most in line with expectations. WSS values in other methods are not the best, but the results are still acceptable.

**4.3. Characteristics of OSI of Bypass Grafts.** OSI is used to reveal the degree of WSS vector direction change which can help us understand the oscillation information of blood flow [36]. Simulation studies have pointed out that low OSI may reduce the risk of graft failure [23, 24]. *In vitro* studies using bovine endothelial cells revealed that slowly oscillating (1 Hz) shear stress (equal to the upper limit of OSI of 0.5) induced the expression of monocyte chemoattractant protein-1 and increased binding of monocytes to the endothelium which indicate high OSI has the potential to increase the risk of vascular failures [37]. Smaller OSI means more stable blood flow and better guarantee of blood vessel stability. The simulation results of this study showed that OSI in all SVGs (with and without CAA ligation) is quite small and close to 0. OSI in LIMA (without CAA ligation, models 5–7) is quite high with an average value of 0.367. OSI in LIMA (with CAA ligation, model 8) is quite small which is close to 0. The results of different surgical methods are significantly different. OSI levels are two orders of magnitude lower in the

SVG with and without CAA ligation than LIMA without CAA ligation. LIMA with CAA ligation has similar condition as the SVG. Therefore, the simulation results of the SVG and LIMA with CAA ligation are in line with the target. LIMA without CAA ligation is at risk due to too high risk of failure.

**4.4. Summing-Up and Deduction.** As mentioned above, we extracted and analyzed three key hemodynamic parameters including mass flow rate, WSS, and OSI. The results show that high mass flow rate, relatively low WSS, and low OSI (almost 0) are the main features for the SVG with and without CAA ligation. Low mass flow rate, low WSS, and high OSI are the main features for the LIMA without CAA ligation. In contrast, high mass flow rate, high WSS, and low OSI (almost 0) are the main features for the LIMA with CAA ligation. The results show that changes in CAA diameter have no significant effect on hemodynamic features of the same category. According to previous research, achieving high mass flow, high WSS, and low OSI to reduce adverse hemodynamic factors is important, in order to reduce the risk of surgical failure and prolonging long-term effectiveness of the graft. Among the several surgical methods in this study, LIMA with CAA ligation has the optimal hemodynamic parameters compared with other models. SVG with and without CAA ligation have similar hemodynamic conditions which are all anticipated apart from the relatively low WSS. LIMA without CAA ligation has negative hemodynamic parameter results that are not expected. Based on our hemodynamic results, we deduced that surgical outcomes may have nothing to do with CAA diameter. LIMA with CAA ligation may have the best surgical

outcomes, and SVG (both with and without CAA ligation) surgical outcomes are better than that of LIMA without CAA ligation. LIMA without CAA ligation may not be desirable. When a coronary artery is ligated, it means that the blood supply to this branch is completely cut off. The blood supply at this time can only rely on the bypass grafts. Once the bypass grafts have problems, the blood vessels themselves cannot perform compensatory adjustments in time which originally can be performed without ligation. Therefore, even though CABGs with CAA ligation may have better surgical outcomes, we need to be very careful with this surgical approach due to other potential risks.

**4.5. Limitations of the Work.** An important limitation of this study is the number of patients. The model in this study was only based on one patient's data, so the results of the study may be special and accidental. The second limitation is the assumption of rigid arterial walls in the simulation, which does not account for the significant motion and curvature changes of the coronary arteries and bypass grafts during the cardiac cycle. In addition, rigid arterial walls mean that the effect of material deformation on blood flow is ignored. This problem can be solved by adopting a fluid-solid interaction (FSI) method in the future. An additional limitation is limited data on flow split percentages to the left and right coronary arteries and the relationship between the intramyocardial pressures to the left and right ventricular pressures, and these are possible minor sources of error during building the LPM.

## 5. Conclusion

In this paper, the hemodynamic differences between two different surgical methods (types of CABGs, whether CAA ligated or not) and the effect of CAA diameter on the hemodynamics of different surgical methods were compared by using CFD simulation. A hemodynamic-based evaluation of these procedures was performed. The results suggest that LIMA with CAA ligation has the best hemodynamic characteristics, including high mass flow rate, high WSS, and low OSI. Hemodynamic characteristics of the SVG (with and without CAA ligation) are better than those of the LIMAs without CAA ligation including high mass flow rate, WSS, and low OSI. Significant reverse flow that can increase the risk of graft failure was found in LIMAs without CAA ligation during systole. The change of CAA diameter has basically no effect on the hemodynamics of the same bypass graft. Based on the previous hemodynamic research, we deduced that the outcomes of ligation surgery may have more benefits than those of nonligation surgery. SVG may be a better choice than LIMA when CAA is not ligated. This preliminary result warrants extensive research involving a larger aneurysm database for further analysis.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

This research was supported by the National Natural Science Foundation of China (nos. 11832003 and 11772016), and JSPS KAKENHI (no. JP18K18355). This research was partially supported by the creation of a development platform for implantable/wearable medical devices by a novel physiological data integration system of the Program on Open Innovation Platform with Enterprises, Research Institute and Academia (OPERA) from the Japan Science and Technology Agency (JST).

## References

- [1] M. Syed and M. Lesch, "Coronary artery aneurysm: a review," *Progress in cardiovascular diseases*, vol. 40, no. 1, pp. 77–84, 1997.
- [2] S. Abou Sherif, O. Ozden Tok, Ö. Taşköylü, O. Goktekin, and I. D. Kilib, "Coronary artery aneurysms: a review of the epidemiology, pathophysiology, diagnosis, and treatment," *Frontiers in Cardiovascular Medicine*, vol. 4, 2017.
- [3] C. Indolfi, F. Achille, G. Tagliamonte, C. Spaccarotella, A. Mongiardo, and A. Ferraro, "Polytetrafluoroethylene stent deployment for a left anterior descending coronary aneurysm complicated by late acute anterior myocardial infarction," *Circulation*, vol. 112, no. 5, pp. 70–71, 2005.
- [4] K. G. Friedman, K. Gauvreau, A. Hamaoka-Okamoto et al., "Coronary artery aneurysms in kawasaki disease: risk factors for progressive disease and adverse cardiac events in the US population," *Journal of the American Heart Association*, vol. 5, no. 9, 2016.
- [5] S. K. Singh, T. Goyal, R. Sethi et al., "Surgical treatment for coronary artery aneurysm: a single-centre experience," *Interactive Cardiovascular and Thoracic Surgery*, vol. 17, no. 4, pp. 632–636, 2013.
- [6] E. Beckmann, S. Rustum, S. Marquardt et al., "Surgical treatment of coronary artery aneurysms," *Journal of Cardiac Surgery*, vol. 32, no. 11, pp. 674–679, 2017.
- [7] J. E. Markis, C. D. Joffe, P. F. Cohn, D. J. Feen, M. V. Herman, and R. Gorlin, "Clinical significance of coronary arterial ectasia," *The American Journal of Cardiology*, vol. 37, no. 2, pp. 217–222, 1976.
- [8] J. Hillebrand, A. Rukosujew, S. Martens, and D. Boese, "Redo operation of recurrent giant coronary artery aneurysm: optimizing surgical strategy," *The Thoracic and Cardiovascular Surgeon Reports*, vol. 5, no. 1, pp. 57–59, 2016.
- [9] A. Dolapoglu and D. A. Ott, "Giant right coronary artery aneurysm associated with a fistula draining into the superior vena cava," *Texas Heart Institute Journal*, vol. 43, no. 4, pp. 360–362, 2016.
- [10] A. Naraen, P. Reddy, C. Notarstefano, and M. Kudavali, "Giant coronary artery aneurysm in a middle-aged woman," *The Annals of Thoracic Surgery*, vol. 103, pp. 313–315, 2017.
- [11] S. Komoda, T. Komoda, E. Ivanitskaia-Kuehn, S. Dreyse, M. Pasic, and R. Hetzer, "Giant aneurysm of the right coronary artery and fistula to the coronary sinus," *General Thoracic and Cardiovascular Surgery*, vol. 58, no. 2, pp. 78–81, 2010.

- [12] M. Gaudino, D. Taggart, H. Suma, J. D. Puskas, F. Crea, and M. Massetti, "The choice of conduits in coronary artery bypass surgery," *Journal of the American College of Cardiology*, vol. 66, no. 15, pp. 1729–1737, 2015.
- [13] J.-H. Wi, H.-C. Joo, Y.-N. Youn, S.-W. Song, T. H. Kim, and K.-J. Yoo, "Comparison of radial artery and saphenous vein composite  $\gamma$  grafts during off-pump coronary artery bypass," *The Korean Journal of Thoracic and Cardiovascular Surgery*, vol. 46, no. 4, pp. 265–273, 2013.
- [14] H. A. Al-Sabti, A. Al Kindi, K. Al-Rasadi, Y. Banerjee, K. Al-Hashmi, and A. Al-Hinai, "Saphenous vein graft vs. radial artery graft searching for the best second coronary artery bypass graft," *Journal of the Saudi Heart Association*, vol. 25, no. 4, pp. 247–254, 2013.
- [15] N. G. Baikoussis, N. A. Papakonstantinou, and E. Apostolakis, "Radial artery as graft for coronary artery bypass surgery: advantages and disadvantages for its usage focused on structural and biological characteristics," *Journal of Cardiology*, vol. 63, no. 5, pp. 321–328, 2014.
- [16] H. Anzai, T. Watanabe, X. Han et al., "Endothelial cell distributions and migration under conditions of flow shear stress around a stent wire," *Technology and Healthcare*, 2019, in Press.
- [17] H. Anzai, B. Chopard, and M. Ohta, "Combinational optimization of strut placement for intracranial stent using a realistic aneurysm," *Journal of Flow Control, Measurement & Visualization*, vol. 2, no. 2, pp. 67–77, 2014.
- [18] M. Zhang, Y. Li, X. Zhao et al., "Haemodynamic effects of stent diameter and compaction ratio on flow-diversion treatment of intracranial aneurysms: a numerical study of a successful and an unsuccessful case," *Journal of Biomechanics*, vol. 58, pp. 179–186, 2017.
- [19] D. K. H. Leong, V. Ashok, A. Nishkantha, Y. H. Shan, and E. K. W. Sim, "Transit-time flow measurement is essential in coronary artery bypass grafting," *The Annals of Thoracic Surgery*, vol. 79, no. 3, pp. 854–857, 2005.
- [20] K. Honda, Y. Okamura, Y. Nishimura et al., "Graft flow assessment using a transit time flow meter in fractional flow reserve-guided coronary artery bypass surgery," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 149, no. 6, pp. 1622–1628, 2015.
- [21] A. M. Plata, S. J. Sherwin, and R. Krams, "Endothelial nitric oxide production and transport in flow chambers: the importance of convection," *Annals of Biomedical Engineering*, vol. 38, no. 9, pp. 2805–2816, 2010.
- [22] R. Kaunas, H. Kang, and K. J. Bayless, "Synergistic regulation of angiogenic sprouting by biochemical factors and wall shear stress," *Cellular and Molecular Bioengineering*, vol. 4, no. 4, pp. 547–559, 2011.
- [23] L.-D. Jou and M. E. Mawad, "Hemodynamic effect of neuroform stent on intimal hyperplasia and thrombus formation in a carotid aneurysm," *Medical Engineering & Physics*, vol. 33, no. 5, pp. 573–580, 2011.
- [24] T. Meirson, E. Orion, C. Di Mario et al., "Flow patterns in externally stented saphenous vein grafts and development of intimal hyperplasia," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 150, no. 4, pp. 871–879, 2015.
- [25] H. J. Kim, I. E. Vignon-Clementel, J. S. Coogan, C. A. Figueroa, K. E. Jansen, and C. A. Taylor, "Patient-specific modeling of blood flow and pressure in human coronary arteries," *Annals of Biomedical Engineering*, vol. 38, no. 10, pp. 3195–3209, 2010.
- [26] R. Agujetas, M. R. González-Fernández, J. M. Nogales-Asensio, and J. M. Montanero, "Numerical analysis of the pressure drop across highly-eccentric coronary stenoses: application to the calculation of the fractional flow reserve," *Biomedical Engineering Online*, vol. 17, pp. 1–22, 2018.
- [27] C. A. Taylor, T. A. Fonte, and J. K. Min, "Computational fluid dynamics applied to cardiac computed tomography for noninvasive quantification of fractional flow reserve," *Journal of the American College of Cardiology*, vol. 61, no. 22, pp. 2233–2241, 2013.
- [28] I. E. Vignon-Clementel, C. A. Figueroa, K. E. Jansen, and C. A. Taylor, "Outflow boundary conditions for 3D simulations of non-periodic blood flow and pressure fields in deformable arteries," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 13, no. 5, pp. 625–640, 2010.
- [29] D. Katritsis, L. Kaiktsis, A. Chaniotis, J. Pantos, E. P. Efstathopoulos, and V. Marmarelis, "Wall shear stress: theoretical considerations and methods of measurement," *Progress in Cardiovascular Diseases*, vol. 49, no. 5, pp. 307–329, 2007.
- [30] F. P. Salvucci, C. A. Perazzob, S. Salles, J. G. Barrac, and R. L. Armentano, "Influence of inlet conditions in wall shear stress distributions of left coronary arteries in patient-specific simulations," in *Mecánica Computacional*, E. Dvorkin, M. Goldschmit, and M. Storti, Eds., vol. XXIX, pp. 5953–5960, Asociacion Argentina, Buenos Aires, Argentina, 2010.
- [31] M. Piccinelli, A. Veneziani, D. A. Steinman, A. Remuzzi, and L. Antiga, "A framework for geometric analysis of vascular structures: application to cerebral aneurysms," *IEEE Transactions on Medical Imaging*, vol. 28, no. 8, pp. 1141–1155, 2009.
- [32] G. D'Ancona, H. L. Karamanoukian, M. Ricci, S. Schmid, J. Bergsland, and T. A. Salerno, "Graft revision after transit time flow measurement in off-pump coronary artery bypass grafting," *European Journal of Cardio-Thoracic Surgery*, vol. 17, no. 3, pp. 287–293, 2000.
- [33] M. Kroll, J. Hellums, L. McIntire, A. Schafer, and J. Moake, "Platelets and shear stress," *Blood*, vol. 88, no. 5, pp. 1525–1541, 1996.
- [34] S.-W. Lee, L. Antiga, and D. A. Steinman, "Correlations among indicators of disturbed flow at the normal carotid bifurcation," *Journal of Biomechanical Engineering*, vol. 131, no. 6, Article ID 061013, 2009.
- [35] A. M. Malek, S. L. Alper, and S. Izumo, "Hemodynamic shear stress and its role in atherosclerosis," *JAMA*, vol. 282, no. 21, pp. 2035–2042, 1999.
- [36] A. Ruiz-Soler, F. Kabinejadian, M. A. Slevin, P. J. Bartolo, and A. Keshmiri, "Optimisation of a novel spiral-inducing bypass graft using computational fluid dynamics," *Scientific Reports*, vol. 7, 2017.
- [37] J. Hwang, M. H. Ing, A. Salazar et al., "Pulsatile versus oscillatory shear stress regulates NADPH oxidase subunit expression," *Circulation Research*, vol. 93, no. 12, pp. 1225–1232, 2003.



## Research Article

# Study on the 2D Optimization Simulation of Complex Five-Hole Cutting Blasting under Different Lateral Pressure Coefficients

Jing Gao,<sup>1</sup> Shizhen Xie,<sup>1</sup> Xiantang Zhang <sup>1,2</sup> Hongli Wang,<sup>2</sup> Wenle Gao,<sup>1</sup> and Hongmin Zhou<sup>1</sup>

<sup>1</sup>School of Civil Engineering and Architecture, Shandong University of Science and Technology, Qingdao 266590, China

<sup>2</sup>Shandong Provincial Key Laboratory of Civil Engineering Disaster Prevention and Mitigation, Shandong University of Science and Technology, Qingdao 266590, China

Correspondence should be addressed to Xiantang Zhang; zzxhtm@sduast.edu.cn

Received 4 April 2020; Accepted 12 May 2020; Published 10 June 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Jing Gao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the decrease and depletion of shallow coal resources, the depth of mining is increasing. The mechanism of high crustal stress blasting is not clear, and the effect of crustal stress on blasting effect is obvious. The law of the differential detonation is similar without crustal stress. The crustal stress of rock masses increases linearly with the increase in excavation depth, and the influence of crustal stress on blasting effect is great. In order to study the rock-breaking process of complex differential blasting under deep high crustal stress, the instantaneous detonation of each model and the priority detonation of the central hole are numerically simulated. The evolution law of the blasting crack and the method of optimizing blasting effect by differential blasting and increasing the aperture of the priority detonation hole under high crustal stress are put forward. The authors proposed a study on the 2D optimization simulation of complex five-hole cutting blasting under different lateral pressure coefficients of 400 m and 800 m with software ANSYS/LS-DYNA and analyzed the evolution rules of blasting cracks and lateral pressure coefficients. The results show that setting delayed detonation and hole diameter and optimizing blasthole spacing can optimize the rock-breaking effect under high crustal stress and different lateral pressure coefficients.

## 1. Introduction

With the reduction and depletion of shallow coal resources, deep mining has become an inevitable trend. With the passage of time, the future will be mined towards deeper rock formations. In the current excavation process of the deep rock mass, the drilling and blasting method is still the main method of rock breaking. As the depth of the excavation increases, the influence of the initial crustal stress on the blasting effect becomes more and more important. According to the variation of vertical and horizontal stress with depth in various countries given by Brady and Brown, the vertical stress generally increases linearly with the increase of depth, and the variation law of horizontal stress is more complicated, which makes the blasting excavation of deep rock mass have great difficulties [1, 2]. At present, the research on deep rock mass mining is still in the initial stage,

the design of excavation parameters is still based on the design method of shallow rock mass blasting, and the influence of initial crustal stress is not considered enough, which makes the blasting effect poor, the excavation speed slow, the late support time long, and even have the rock burst and other major disasters. This makes it particularly important to study blasting rock breaking under initial crustal stress, and it is particularly urgent to study the design of blasting parameters and the rapid and safe excavation of rock mass stability under initial crustal stress.

In view of the rock-breaking mechanism of deep rock mass under initial crustal stress, scholars at home and abroad have carried out extensive research on physical experiment and numerical simulation. Based on the stress state equation of the rock mass in the cut area established by Schwarz alternating iteration method and elastic mechanics theory, two-dimensional numerical simulation of

straight cut blasting under different stress fields was carried out, and the maximum tensile stress was generated on the line connecting the hollow hole and the slot hole after the initial crustal stress field and the explosion stress field were combined by the rock mass [2]. The damaged model of rock blasting built on the basis of damaged mechanics analyzed the influence of different lateral pressure coefficients and buried depth on the crack growth rule [3]. The coupling model of Cowper–Symonds hardening model and tension and compression damaged model was embedded in LS-DYNA software for a two-dimensional numerical simulation. It was found that the central hole had an impact on rock damage [4]. Blasting parameter design method is suitable for deep high-stress rock roadway excavation by using numerical simulation and field test [5]. Wang et al. did numerical simulation by creep experimental data, obtained the creep equation of the surrounding rock in the deep soft rock roadway under high crustal stress, and established the quantitative creep relationship between time and stress [6]. He et al. summarized and analyzed the main differences of rock mass mechanical properties between deep mining and shallow mining through experiments [7]. The theoretical calculation model of the fragmentation size distribution containing the damage variable is proposed. Simulations and tests show that the action of the extrusion and collision between motive fragment blocks is a significant factor affecting the average fragmentation size regardless of the type of rock mass, especially for the rock mass with very bad quality [8]. The new model named MAGRM is a group recommendation model based on multiattention. It can well utilize multiattention-based deep neural network structures to achieve accurate group recommendation [9]. Based on LS-DYNA finite element program, NA finite element program, and optimization of cutting blasting in high-speed drivage process of an iron mine roadway, the authors carried out a study about numerical simulation and field test to the cutting blasting method with different blasting holes [10]. In order to eliminate the simplification of the interaction in data processing and improve the generated working efficiency, a method of redundant electrode determination based on the theory of variance is proposed [11]. Simulation of the complex rock-breaking mechanism under crustal stress included the prediction model of carbonation depth of slag high-performance concrete [12, 13]. Based on rock blasting theory, the authors used ANSYS/LS-DYNA to simulate the process of rock fragmentation of double-blasthole blasting with one empty hole at the center position [14]. The crack propagation in the rock under composite stress was studied by using digital image correlation technique [15]. The design and realization of the magnetic suspension controller are discussed, and a nonlinear mathematical model of the magnetic suspension system is built [16]. Based on the theory of the BP neural network and the engineering geological database as the research and development platform, this paper establishes the prediction of geotechnical parameters based on the analysis of the characteristics of geotechnical materials and the distribution of geotechnical sediments and geotechnical

parameters [17]. Local plane wave decomposition method considered from the perspective of inversion is employed to obtain higher-quality slant stack data, and the dynamic focused beam propagator is adopted to control the divergence of beams [18]. To reduce the impact of blasting vibration on the existing tunnel during the excavation of small-distance tunnel blasting, software was used to simulate the blasting vibration of the tunnel in different positions and forms [19].

Under the action of crustal stress, two-dimensional numerical simulation of complex five-hole blasting under different lateral pressure coefficients is carried out using ANSYS/LS-DYNA. Aiming at the differential detonation and increasing the diameter of the central blasthole under high crustal stress, the simulation study is provided for the blasting construction of deep rock mass.

## 2. Theoretical Analysis of Rock State with Different Confining Pressures and Lateral Pressure Coefficients

The drilling of the rock under the condition of initial crustal stress has changed the original stable state of the rock so that the rock around the blasthole produces an instant loosening zone. Since the blasthole is relatively small compared to the roadway, the concentrated stress value around the blasthole is smaller than the rock strength, the surrounding rock is in an elastoplastic state, and the surrounding rock is self-stabilizing. However, the stability of the loosening zone around the blasthole is less than that of the original rock. Similar to the loosening circle theory, in bidirectional confining pressure (side lateral pressure coefficient is equal to 1), vertical stress and horizontal stress are equal, the loosening zone is circular, the rock stability around the blasthole is evenly distributed, and the direction of rock crack propagation after explosion is randomly distributed, which is inhibited by the crustal stress. In the case of bidirectional unequal confining pressure (lateral pressure coefficient is not equal to 1), vertical stress and horizontal stress are not equal, the loosening zone is elliptic, and the long axis of the ellipse is perpendicular to the direction of the maximum crustal stress. Rock around the blasthole is affected by the initial stress field, and the explosion crack is restrained by the crustal stress and mainly expands to the side with large crustal stress. Figures 1–3, respectively, show the effect diagrams when the lateral pressure coefficient is 1, 0.2, and 4 under the initial vertical stress of 10 MPa. As can be seen from the figure, when the lateral pressure coefficient  $\lambda = 1$ , uniform stress concentrations are produced around the rock. When the lateral pressure coefficient  $\lambda = 0.2$  and 4, the tensile stress concentration is generated on the side with larger crustal stress. The maximum tensile stress is generated by the coupling of crustal stress and explosion stress under the action of constant pressure on both sides, this analysis analyzed the cause of the maximum tensile stress and the influence of different side pressures on the maximum tensile stress direction from the initial stress of the blasthole [2].

The initial stress state of the rock has an important effect on the blasting rock. Rock blasting under the initial stress

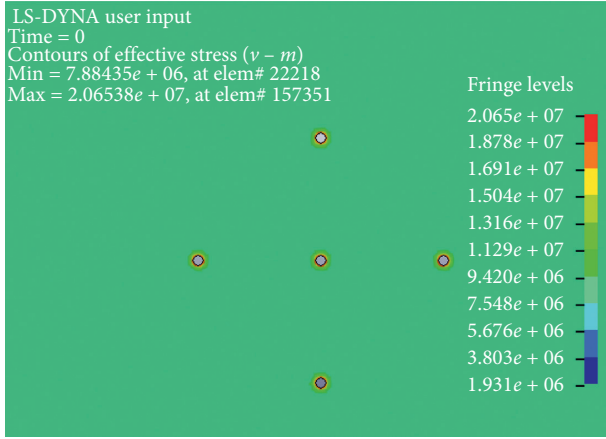


FIGURE 1: The effect diagram when the lateral pressure coefficient is 1.

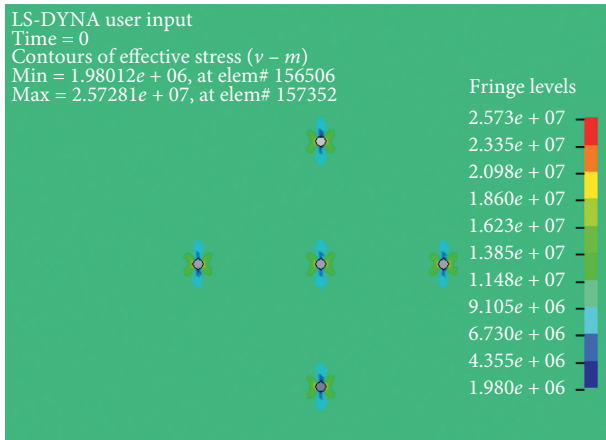


FIGURE 2: The effect diagram when the lateral pressure coefficient is 0.2.

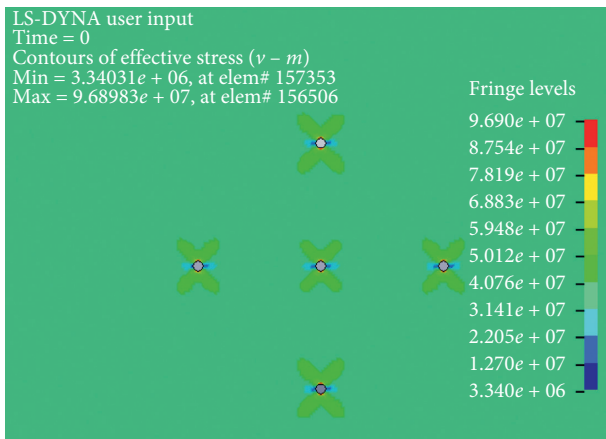


FIGURE 3: The effect diagram when the lateral pressure coefficient is 4.

condition is the effect of the coupling field formed by the coupling of the stress field redistributed in the rock around the blasthole and the explosive stress field after the explosion.

2.1. *Calculation of the Initial Crustal Stress Field of the Rock.* Radial effective stress  $\sigma_{r0}$ , tangential effective stress  $\sigma_{\theta0}$ , and shear stress of the rock  $\tau_{r\theta}$  around the blasthole under initial crustal stress can be calculated according to the elastic theory [20]:

$$\begin{cases} \sigma_{r0} = \frac{\sigma_H + \sigma_h}{2} \left(1 - \frac{R^2}{L^2}\right) + \frac{\sigma_H - \sigma_h}{2} \left(1 + \frac{3R^4}{L^4} - \frac{4R^4}{L^4}\right) \cos 2\theta, \\ \sigma_{\theta0} = \frac{\sigma_H + \sigma_h}{2} \left(1 + \frac{R^2}{L^2}\right) + \frac{\sigma_H - \sigma_h}{2} \left(1 + \frac{3R^4}{L^4}\right) \cos 2\theta, \\ \sigma_{r\theta} = \frac{\sigma_H - \sigma_h}{2} \left(1 - \frac{3R^4}{L^4} + \frac{2R^4}{L^4}\right) \sin 2\theta, \end{cases} \quad (1)$$

where  $\sigma_H$  and  $\sigma_h$  are the maximum and the minimum horizontal crustal stress;  $\theta$  is the angle between the radial direction of the point around the hole and the direction of the maximum horizontal crustal stress;  $R$  is the radius of the blasthole; and  $L$  is the distance from the calculation point to the center of the blasthole.

2.2. *Calculation of the Rock Stress under Explosive Load.* The stress state around the rock is further changed by the reflection and tension of the stress wave on the rock around the blasthole under the explosion stress [21]. The radial stress of an explosion at a certain point in the rock  $\sigma_{r1}$  and the tangential stress of an explosion at a certain point in the rock  $\sigma_{\theta1}$  can be calculated according to the attenuation formula of the stress wave [22]:

$$\begin{cases} \sigma_{r1} = p_0 \left(\frac{L}{R}\right)^{-\alpha}, \\ \sigma_{\theta1} = -\lambda_d \sigma_r, \end{cases} \quad (2)$$

where  $p_0$  is the initial pressure acting on the hole wall after the explosion of the explosive; the value of  $\alpha$ , stress wave attenuation coefficient, is  $2 - (\mu_d/1 - \mu_d)$ ;  $\mu_d$  is the dynamic Poisson ratio and the relationship between it and the static Poisson ratio is  $\mu_d = 0.8\mu$ ; and  $\lambda_d$  is the dynamic lateral stress coefficient, and its value is  $\lambda_d = (\mu_d/1 - \mu_d)$ .

Therefore, the tangential stress after superposition is  $\sigma_r = \sigma_{r0} + \sigma_{r1}$ . That is, when the tangential stress in the rock is greater than the ultimate tensile strength of the rock under the combined action of the initial stress and the explosion stress, the rock will produce cracks, and then failure occurs.

### 3. Establishment of a Numerical Model for Rock Blasting

3.1. *Rock Constitutive Model and Material Parameters.* In order to observe the change of the rock damaged vertically and horizontally under the initial crustal stress field, the theoretical model of complex five-hole cutting blasting is selected under the condition of RHT rock constitutive, the fixed constraint is applied to the bottom of the model, the surrounding compressive stress is applied in the other

three directions, and the initial crustal stress field is simulated by using the implicit-display order of DYNA [23]. Because the law mainly analyzes and summarizes the damage degree of the rock, it does not involve the formation process of the groove cavity, and it directly simulates the damage evolution process of rock profile with a thickness of 0.01 m. When the computer carries out the numerical simulation, the mesh quality affects the accuracy of the numerical simulation, and the size of the mesh division and the number of the quantity affect the accuracy of the simulation. The model size is  $5\text{ m} \times 4\text{ m} \times 0.01\text{ m}$ , wherein 5 blastholes with a radius of 20 mm are arranged, and the blasthole spacing is 0.5 m. Because the aperture size of the blasthole is small relative to the model size and the interface size does not change vertically with the model, the simulation can be reduced to a plane strain problem calculated by a single layer mesh, and the model is shown in Figure 4.

The explosive material model adopts \*MAT\_HIGH\_EXPLOSIVE\_BURN, using the JWL state equation, and the specific explosive parameters are shown in Table 1 [24]. Rock mechanics parameters are shown in Table 2 [25].

The JWL equation of the state defines that the pressure is

$$p = A \left( 1 - \frac{w}{R_1 V} \right) e^{-R_1 V} + B \left( 1 - \frac{w}{R_2 V} \right) e^{-R_2 V} + \frac{wE}{V}, \quad (3)$$

where  $p$  is the pressure of the detonation product;  $V$  is the relative volume of the detonation product;  $E$  is the initial internal energy density of the detonation product; and  $A$ ,  $B$ ,  $R_1$ ,  $R_2$ , and  $w$  are material parameters.

The material for the cell is \*MAT\_NULL. Defining the mass, the modulus of elasticity, Poisson's ratio, and so on, to define the material. The air material model uses MAT\_NULL air material model, its state equation is defined by the keyword EOS\_LINEAR\_POLYNOMIAL, and the state equation is as follows:

$$p = C_0 + C_1 \mu + C_2 \mu^2 + C_3 \mu^3 + (C_4 + C_5 \mu + C_6 \mu^2) E, \quad (4)$$

where  $C_i$  is the equation parameter and  $E$  is the unit initial internal energy of relative volume  $\mu = (1/V) - 1$ ,  $V$  is the relative volume. Parameters of air material and state equations are shown in Table 3.

### 3.2. Loading Scheme of Crustal Stress with Different Lateral Pressure Coefficients in Blasting Damage

- (1) When the lateral pressure coefficient is  $\lambda = 1$ , model A:  $\sigma_z$  takes 10 MPa and corresponds to 400 m buried depth; model B:  $\sigma_z$  takes 20 MPa and corresponds to 800 m buried depth
- (2) When the lateral pressure coefficient is  $\lambda = 0.2$ , model C:  $\sigma_z$  takes 10 MPa,  $\sigma_x$  to 2 MPa, and model D:  $\sigma_z$  takes 20 MPa,  $\sigma_x$  to 4 MPa
- (3) When the lateral pressure coefficient is  $\lambda = 4$ , model E:  $\sigma_z$  takes 10 MPa,  $\sigma_x$  to 40 MPa, and model F:  $\sigma_z$  takes 20 MPa,  $\sigma_x$  to 80 MPa

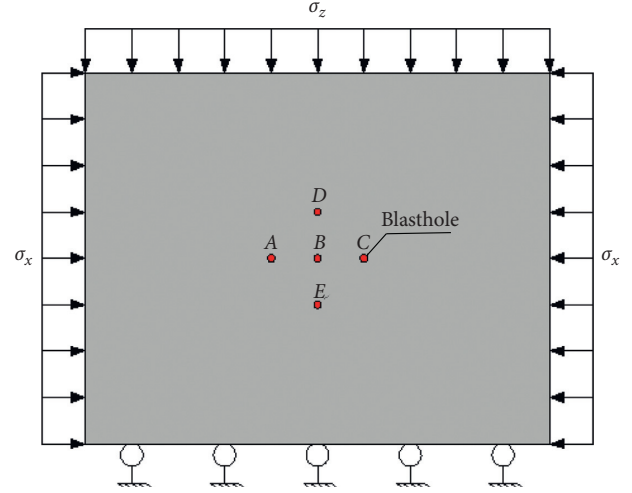


FIGURE 4: Numerical model under a single layer mesh.

TABLE 1: Establishment of the numerical model for blasting damage.

| Density<br>$\text{kg}\cdot\text{m}^{-3}$ | Blast<br>speed<br>$\text{m}\cdot\text{s}^{-1}$ | CJ<br>pressure<br>GPa | JWL state equation parameters |            |       |       |          |            |
|--|--|-----------------------|-------------------------------|------------|-------|-------|----------|------------|
|  |  |                       | A<br>(GPa)                    | B<br>(GPa) | $R_1$ | $R_2$ | $\omega$ | E<br>(GPa) |
| 1,300                                    | 6,000  | 3.240                 | 220                           | 0.2        | 4.5   | 1.1   | 0.35     | 4.2        |

## 4. Numerical Simulation Results and Analysis

In order to study the rock-breaking process of differential blasting under deep high crustal stress, the instantaneous detonation of each model and the prior detonation of the central hole are numerically simulated. The differential blasting initiation model sets the central hole to detonate first, and the remaining holes are set 50  $\mu\text{s}$  time delay and compared under different lateral pressure coefficients; the cloud map is output and analyzed after the computing cluster.

**4.1. Rock Damage Evolution Process with Lateral Pressure Coefficient Equal to 1.** Because the lateral pressure coefficient is equal to 1, the pressure around the hole is evenly distributed; in the early stage of the explosion, the explosion shock wave first acts on the rock mass around the hole so that the rock mass around the blasthole first produces damage. Before 50  $\mu\text{s}$ , the range of the rock damage around the blasthole under the same crustal stress model was similar, the explosion crushing area was basically molded, and the stress wave was not enough to produce direct damage to the rock mass while resisting the crustal stress and began to continue along the damaged area that had been formed. The damage range of the rock is comparable under 10 MPa and 20 MPa crustal stress, and there is no great difference. After 50  $\mu\text{s}$ , with the gradual attenuation of the stress wave, the inhibition of the crustal stress on the blasting effect began to show the following: due to the gradual attenuation of the stress wave, the influence of crustal stress on blasting effect is more and more in the later stage of

TABLE 2: Numerical model rock-specific parameters.

| Parameter                                | Value | Parameter   | Value    |
|--|-------|---|----------|
| Mass density RO ( $\text{kg/m}^3$ )      | 2,660 | Porosity exponent NP                              | 3.0      |
| Initial porosity ALPHA                   | 0     | Reference compressive strain rate EOC             | $3.E-5$  |
| Crush pressure PEL (MPa)                 | 125   | Reference tensile strain rate EOT                 | $3.E-6$  |
| Compaction pressure PCO (GPa)            | 6.0   | Break compressive strain rate EC                  | $3.E+25$ |
| Hugoniot polynomial coefficient A1 (GPa) | 25.7  | Break tensile strain rate ET                      | $3.E+25$ |
| Hugoniot polynomial coefficient A2 (GPa) | 37.84 | Compressive strain rate dependence exponent BETAC | 0.026    |
| Hugoniot polynomial coefficient A3 (GPa) | 21.29 | Tensile strain rate dependence exponent BETAT     | 0.007    |
| Parameter for polynomial EOS B0          | 1.22  | Volumetric plastic strain fraction in tension PTF | 0.001    |
| Parameter for polynomial EOS B1          | 1.22  | Compressive yield surface parameter GC*           | 0.53     |
| Parameter for polynomial EOS T1 (GPa)    | 25.7  | Tensile yield surface parameter GT*               | 0.7      |
| Parameter for polynomial EOS T2          | 0.0   | Erosion plastic strain EPSF                       | 2.0      |
| Elastic shear modulus SHEAR (GPa)        | 21.9  | Shear modulus reduction factor XI                 | 0.5      |
| Compressive strength FC (MPa)            | 167.8 | Damage parameter D1                               | 0.04     |
| Relative tensile strength FT*            | 0.04  | Damage parameter D2                               | 1.0      |
| Relative shear strength FS*              | 0.21  | Minimum damaged residual strain EPM               | 0.015    |
| Failure surface parameter A              | 2.44  | Residual surface parameter AF                     | 0.25     |
| Failure surface parameter N              | 0.76  | Residual surface parameter AN                     | 0.62     |
| Lode angle dependence factor Q0          | 0.68  | Gruneisen gamma GAMMA                             | 0.0      |
| Lode angle dependence factor B           | 0.05  |   |          |

TABLE 3: Parameters of air material and state equations.

| Density ( $\text{g}\cdot\text{cm}^3$ ) | $C_0$ | $C_1$ | $C_2$ | $C_3$ | $C_4$ | $C_5$ | $C_6$ | $E$ (GPa) |
|--|-------|-------|-------|-------|-------|-------|-------|-----------|
| $1.29E-5$                              | 0     | 0     | 0     | 0     | 0.4   | 0.4   | 0     | 0.025     |

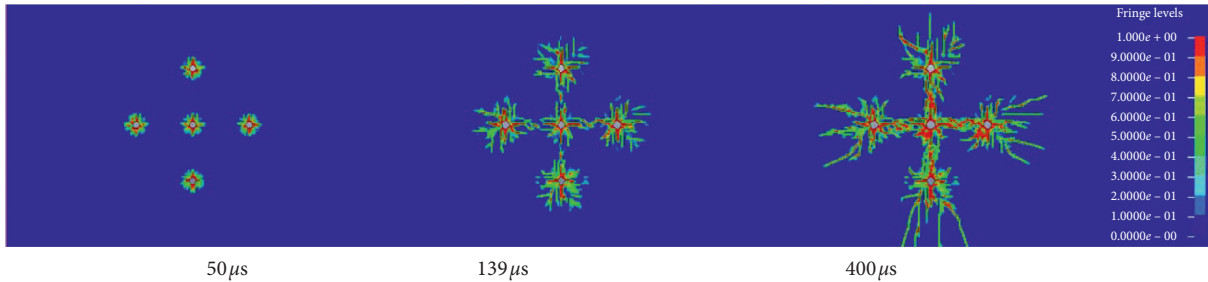


FIGURE 5: Damage evolution process diagram of model A.

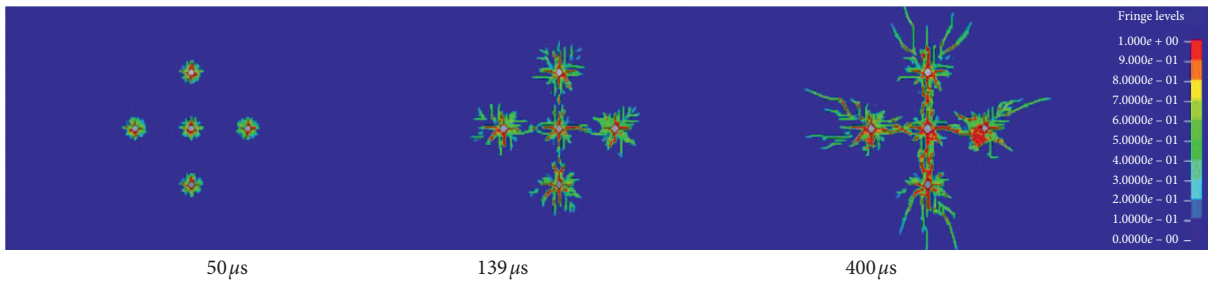


FIGURE 6: Damage evolution process diagram of model B.

explosion, and it can be seen from the final damage cloud map that the damage range and damaged degree of model A are obviously higher than those of model B; the greater the initial crustal stress, the greater the attenuation speed of the

wave. Figures 5 and 6, respectively, show the damage evolution process of models A and B.

Because the lateral pressure coefficient is equal to 1, the evolution of differential blasting damage is also inhibited by

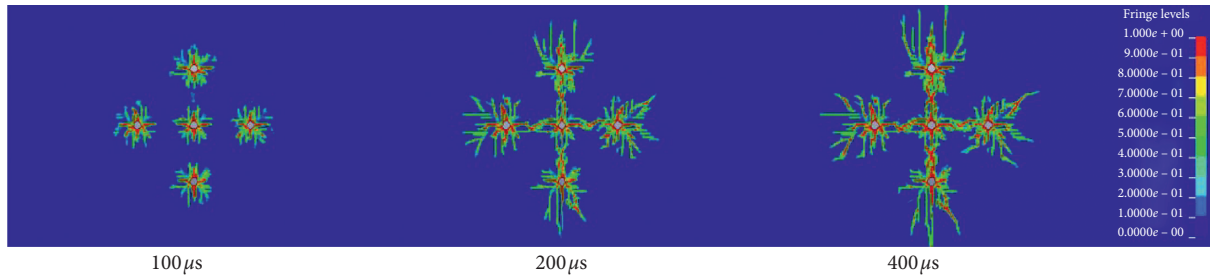


FIGURE 7: Damage evolution process diagram of model C.

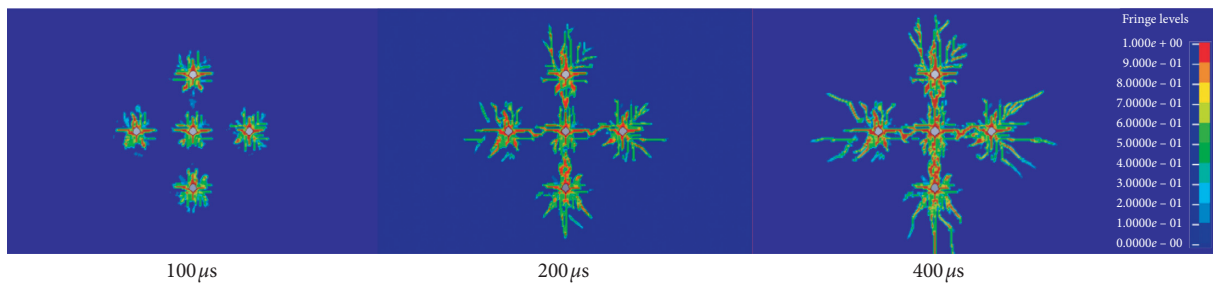


FIGURE 8: Damage evolution process diagram of model D.

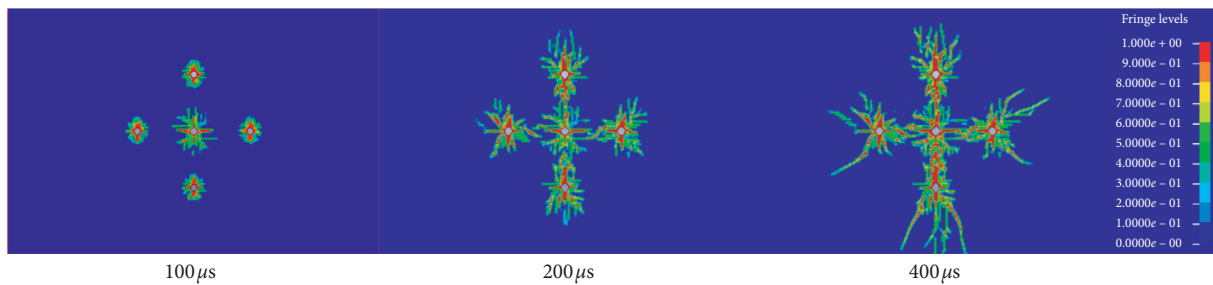


FIGURE 9: Differential initiation damage evolution process diagram of model D.

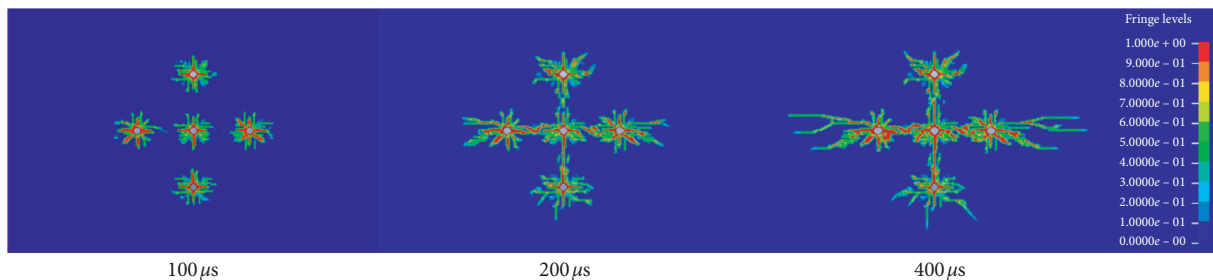
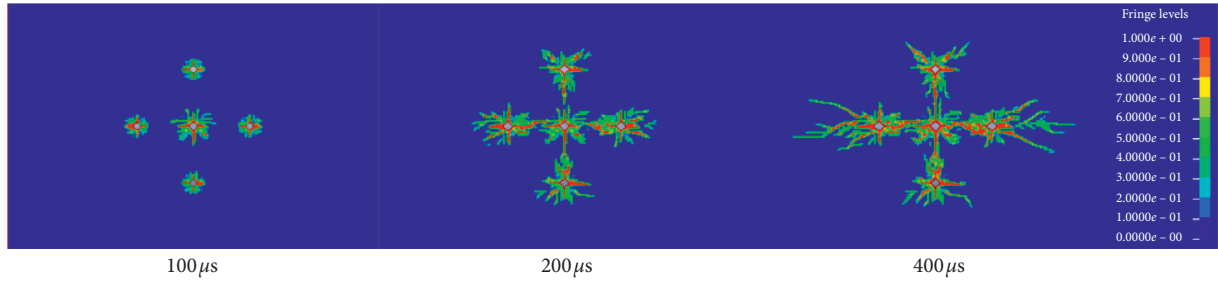
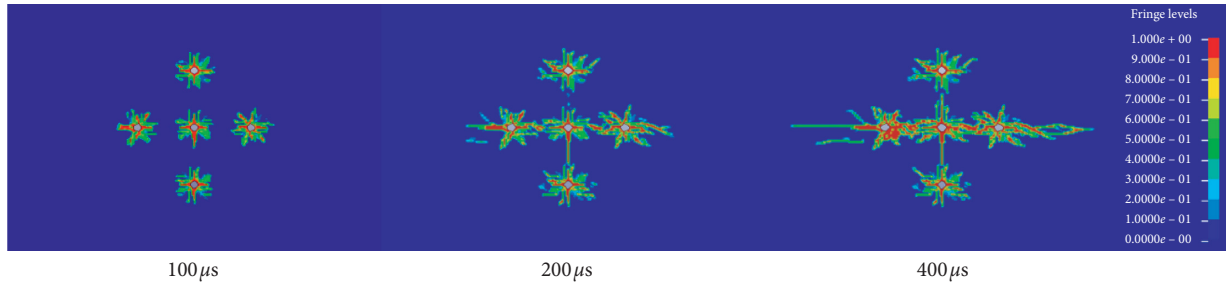
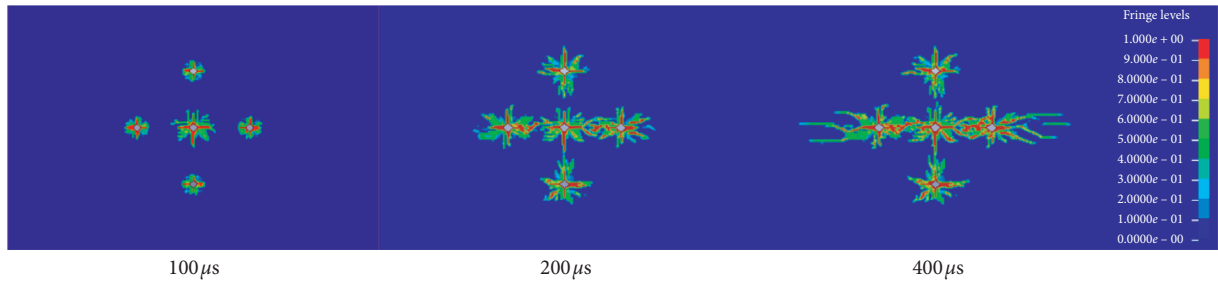
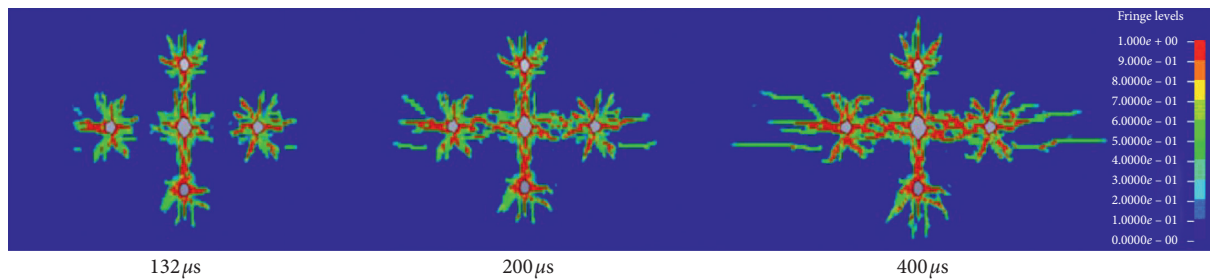


FIGURE 10: Instantaneous detonation damage evolution process cloud map of model E.

the crustal stress, but its evolution law is similar to that of noncrustal stress, so it is no longer described.

*4.2. Rock Damage Evolution Process with Lateral Pressure Coefficient Equal to 0.2.* By analyzing the damage evolution process with a lateral pressure coefficient of 0.2, it can be found that when the lateral pressure coefficient is 0.2, the

initial crustal stress around the hole is smaller when the bidirectional crustal stress is equal, so the inhibition of the initial crustal stress on the rock damage evolution is relatively small at this time. From the damage cloud diagram, it can be seen that because the crustal stress in the vertical direction is greater than the crustal stress in the horizontal direction, the rock damage mainly develops in the vertical direction, the damage range of the hole connection in the

FIGURE 11: Differential initiation damage evolution process diagram of model *E*.FIGURE 12: Instantaneous detonation damage evolution process cloud map of model *F*.FIGURE 13: Differential initiation damage evolution process diagram of model *F*.FIGURE 14: Instantaneous detonation damage evolution process cloud map of model *G*.

vertical direction is larger than the horizontal direction, and the guiding effect of the crustal stress on the crack propagation begins to show. Compared with models *C* and *D*, although the side pressure coefficient is the same, the gap between  $\sigma_z$  and  $\sigma_x$  of model *D* is larger, the guidance effect of model *D* on damage evolution is more obvious, and the damage evolution speed and damage range in the vertical direction are larger than those of model *C*; the gap between  $\sigma_z$  and  $\sigma_x$  in model *C* is smaller, the damage evolution

process in the horizontal and vertical direction is slightly different, but the damage range in the direction of the final hole connection is not much different. It can be seen that the guiding effect and inhibition of crustal stress on damage evolution are not linear. The damage evolution process for models *C* and *D* is shown in Figures 7 and 8.

Because the blasting effect of model *D* in the horizontal direction is not ideal, the central hole is set to detonate first, and the other holes are set to  $50 \mu\text{s}$  time-delay detonation.

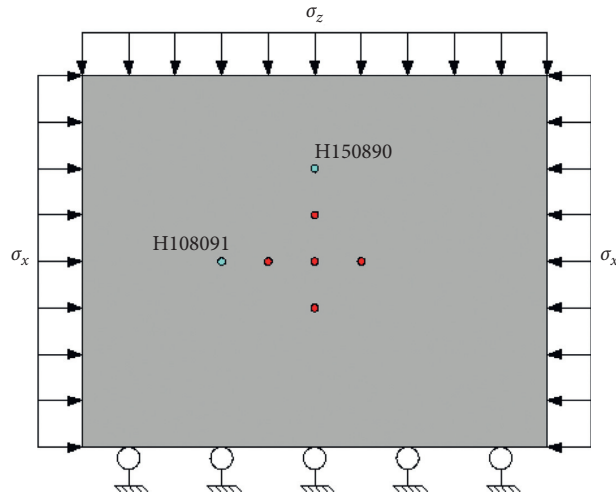
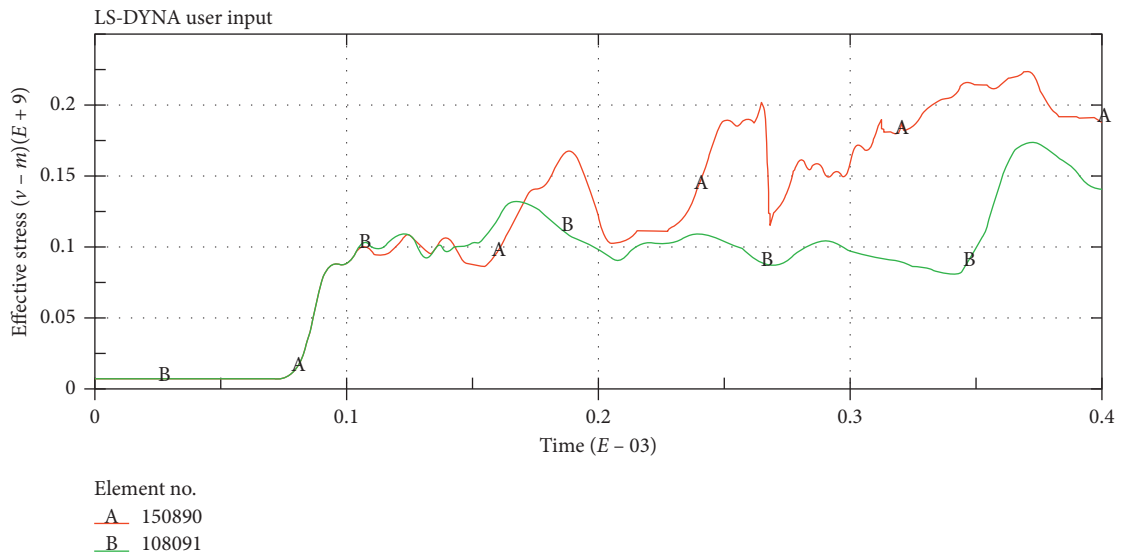
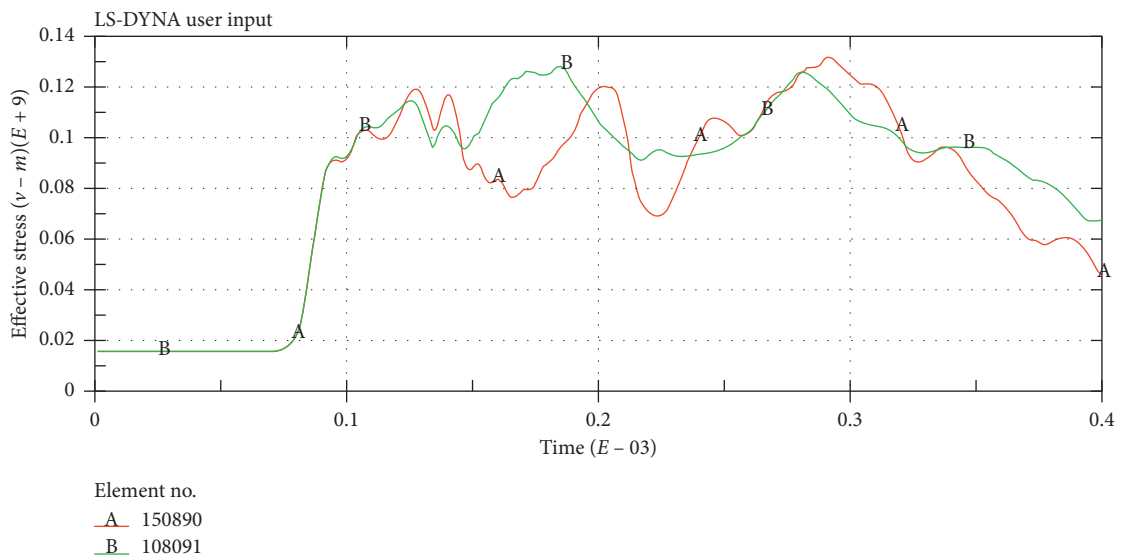


FIGURE 15: Selection diagram of observation points.



(a)



(b)

FIGURE 16: Equivalent stress diagram of observation point. (a) Two-way confining pressure is 10 MPa. (b) Two-way confining pressure is 20 MPa.



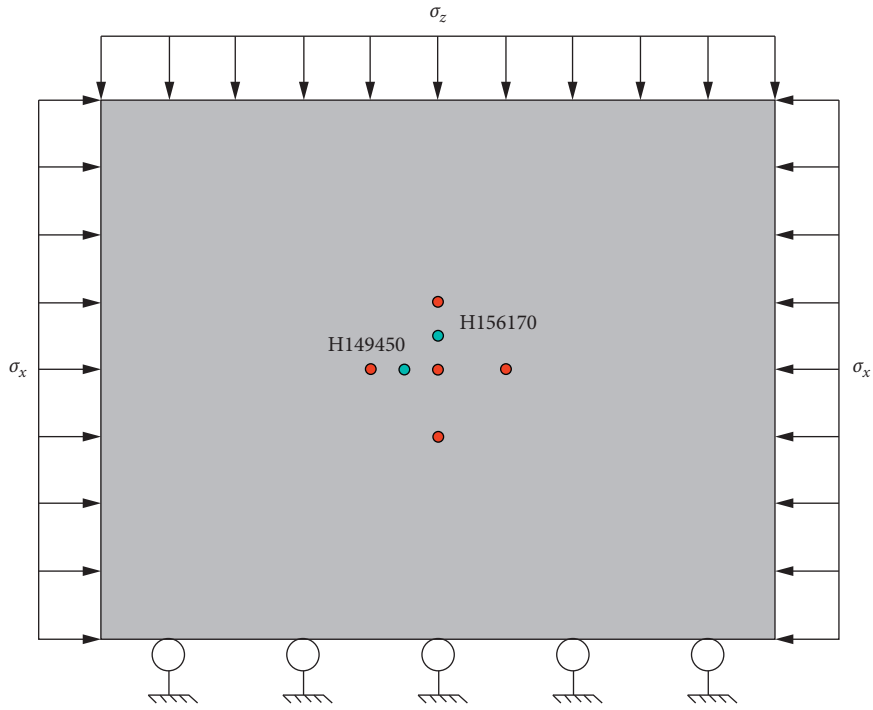
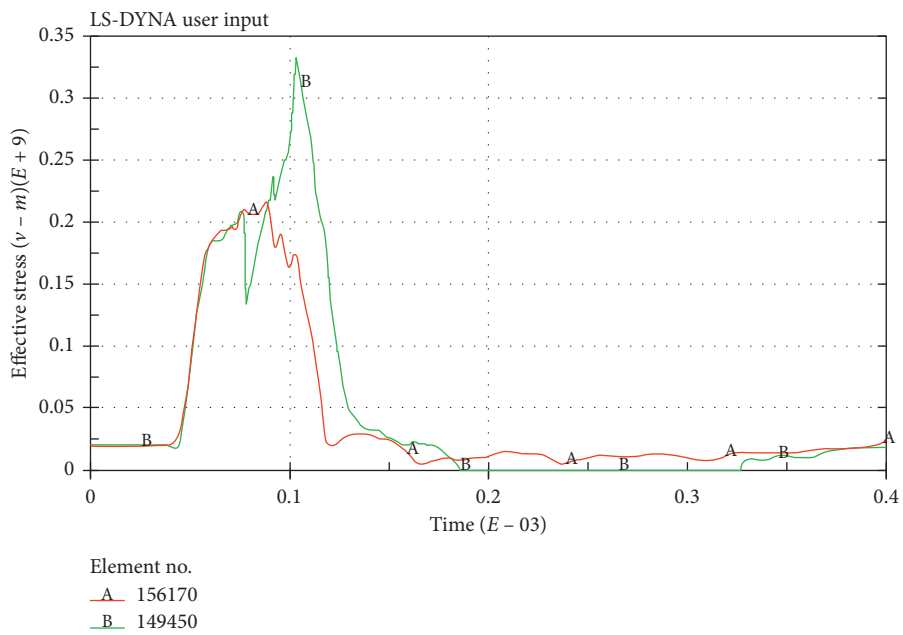
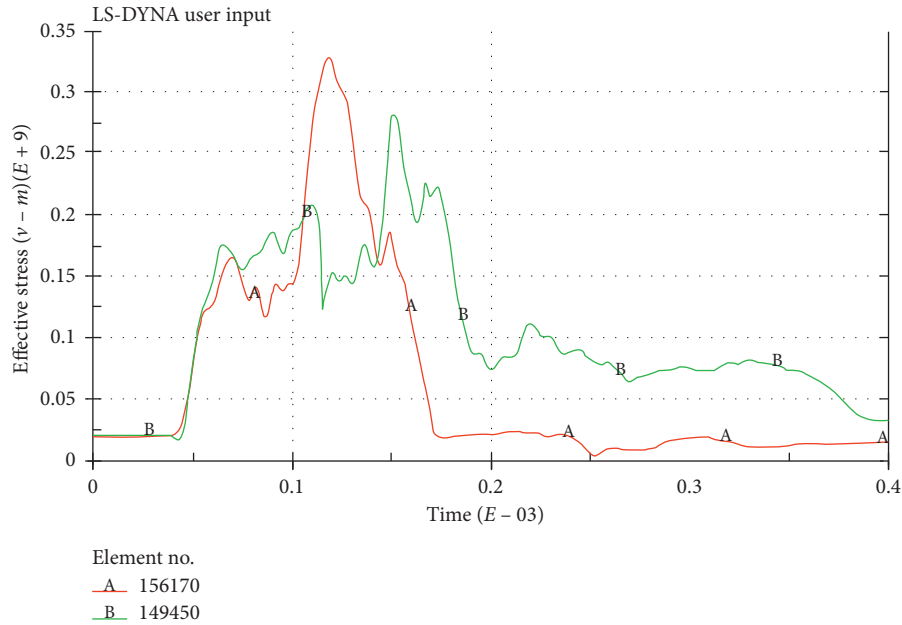


FIGURE 17: Model *D* selects unit location.



(a)

FIGURE 18: Continued.



(b)

FIGURE 18: Comparison of the effective stress time-history curve of the selected element in model *D*. (a) Model *D* time-history curve of effective stress of multichoice element differential blasting. (b) Model *D* time-history curve of effective stress of multichoice element instantaneous blasting.

The damage evolution diagram of model *D* is shown in Figure 9. Compared with the instantaneous detonation, the damage range of the rock is obviously larger than that of instantaneous detonation, the guiding effect of high crustal stress on damage evolution is weakened, the range of horizontal damage is obviously increased, the damage zone on the outside of the vertical hole is shortened, and it is beneficial to the measures such as late support.

**4.3. Rock Damage Evolution Process with Lateral Pressure Coefficient Equal to 4.** With the increase of crustal stress, the inhibitory effect on rock damage evolution increases gradually, while the damaged area is further concentrated in the maximum crustal stress direction, the damage evolution speed in the direction of the maximum crustal stress is accelerated, the region of damage on the outside of the hole is increased, and rock blasting under the initial crustal stress is more difficult. The optimization effect of differential detonation on blasting effect is weakened, which mainly promotes rock damage of the intersection position of the stress wave, and the coupling field between stress wave and crustal stress beyond the limit tensile strength of the rock is further reduced and concentrated. The instantaneous and differential damage evolution diagram of models *E* and *F* with lateral pressure coefficient equal to 4 is, respectively, shown in Figures 10–13.

In model *F*, due to the large difference between  $\sigma_z$  and  $\sigma_x$ , the damage area in the vertical direction is not penetrated. At this time, in the case of differential blasting, it can be considered to increase the diameter of the central hole and shorten the hole spacing on the side with smaller crustal

stress. The model increases the diameter of the central hole to 30 mm, and the blasthole spacing decreases to 0.3 m for numerical simulation, defined as model *G*. The simulation results show that the vertical direction damage is already running through at 132  $\mu$ s, and the damage degree and damage area are still increasing, the final hole is completely running through, increasing the utilization rate of the hole, the difference of damage range in two directions decreases, and there will be no situation similar to that of model *F* blasting. The damage evolution process for model *G* is shown in Figure 14.

**4.4. Simulation Analysis of Five-Hole Cutting Blasting Results.** From the above model and simulation improvement, it can be seen that differential blasting plays a certain role in optimizing the blasting effect under the action of crustal stress, and with the increase of crustal stress, the optimization effect of differential blasting decreases gradually so that reducing the hole spacing in the smaller direction of crustal stress can achieve the optimization effect, and further optimization can increase the diameter of the central hole; in order to meet the blasting demand under the high crustal stress, the utilization rate of the blasthole is improved.

The attenuation law of the stress wave under crustal stress is observed at two observation points H150890 and H108091 in the same position when bidirectional compressive stress is 10 MPa and 20 MPa, respectively, and the selection position is shown in Figure 15.

From the equivalent stress-time curve of the two models, it can be seen that the time for the observation point to reach the first stress peak is similar, around 127  $\mu$ s, which is

because the observation point is close to the hole spacing, the force on the stress wave is not obvious, and the effective stress is around 115 MPa; when the observation point reaches its second peak, at this point, the stress wave begins to overlay, the crustal stress corresponds to the role of the wave which began to produce differences; for the observation points H150890 and H108091 in the confining pressure of 10 MPa, the peak time and size are (189  $\mu$ s, 168 MPa) and (166  $\mu$ s, 132 MPa), and when the confining pressure is 20 MPa, the peak time and size are (205  $\mu$ s, 120 MPa) and (185  $\mu$ s, 126 MPa); combined with the peak data of Figure 16 [26], it can be seen that, with the increase of crustal stress, the influence on the propagation of the stress wave increases, which is reflected in the increase of stress wave velocity and stress size attenuation velocity.

In order to study the impact of differential blasting on rock damage under different lateral pressures, the time-history curves of effective stress of differential blasting and instantaneous blasting in model *D* were selected for comparative analysis. The selected unit is shown in Figure 17.

In Figure 18, the effective stress of model *D* decays rapidly after reaching the peak and decays to the initial crustal stress value at around 112  $\mu$ s, while the differential blasting has experienced stress wave superposition for many times, and the effect time is longer and decays to the effective stress value at around 160  $\mu$ s. It can be seen that the action time of differential blasting is long, which can promote the rock damage. In addition, its peak stress is smaller than that of instantaneous blasting, and the resulting vibration is smaller, which is conducive to the stability of the surrounding rock mass under high crustal stress.

## 5. Conclusions

Using finite element analysis software ANSYS/LS-DYNA, the rock damage of complex five-hole cutting blasting under different lateral pressure coefficients is numerically simulated, and two-dimensional simulation can divide more meshes than three-dimensional simulation at the same time, which can improve the accuracy of simulation when simulating complex blasting. Rock blasting under high crustal stress can be regarded as the damage caused by the coupling position of the rock in the explosion field and the crustal stress field, which is greater than the ultimate tensile strength of the rock.

Under the bidirectional pressure, the crack propagation is inhibited with the increase of crustal stress field, and the evolution law of differential blasting damage is similar to that of noncrustal stress field, which is inhibited by the initial crustal stress field. In the two-way unequal pressure, the direction of damage evolution tends to be on the side of large crustal stress field, and the guiding effect and inhibition of crustal stress field on damage evolution are nonlinear; the priority detonation of the central hole can optimize the blasting effect on the smaller side of the crustal stress field and shorten the damage length of the larger side of the crustal stress field, which decreases with the increase of the crustal stress field; at this time, we can consider increasing the diameter of the priority detonation hole under the

condition of differential blasting. At the same time, shortening the hole spacing on the smaller side of the crustal stress field can further improve the blasting effect. The stress wave velocity and the effective stress attenuation velocity increase with the increase of the crustal stress field.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

This study was supported by the National Natural Science Foundation of China (51874189 and 51709161), the Shandong Provincial Natural Science Foundation of China (ZR2017MEE043), the Visiting Scholar Fund of Shandong University of Science and Technology, and the Project of Shandong University of Science and Technology Graduate Innovation Fund (SDKDYC190241 and SDKDYC190242). The financial supports are gratefully acknowledged.

## References

- [1] E. T. Brown and E. Hoek, "Trends in relationships between measured in-situ stresses and depth," *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*, vol. 15, no. 4, pp. 211–215, 1978.
- [2] Q. T. Li, W. L. Huang, and Z. Y. Wu, "Theoretical study and numerical simulation on rock failure process in cutting by parallel hole under different crustal stress conditions," *Journal of Safety Science and Technology*, vol. 12, no. 11, pp. 57–62, 2016, in Chinese.
- [3] Y. Bai, W. C. Zhu, and C. H. Wei, "Numerical simulation on two-hole blasting under different in-situ stress conditions," *Rock and Soil Mechanics*, vol. 34, no. S1, pp. 466–471, 2013, in Chinese.
- [4] L. Xie, W. Lu, Q. Jiang, and Q. Zhang, "Damaged evolution mechanism of deep rock mass in process of cut blasting," *Journal of Central South University Science and Technology*, vol. 48, no. 5, pp. 1252–1260, 2016, in Chinese.
- [5] H. J. Wang, R. S. Yang, and Q. Li, "Analysis of blasting mechanism for deep rock tunneling and blasting parameters design," *Journal of China Coal Society*, vol. 4, pp. 373–376, 2007, in Chinese.
- [6] Y. Y. Wang, J. Wei, and J. Qi, "Study on prediction for nonlinear creep deformation of deep rocks," *Journal of China Coal Society*, vol. 4, pp. 409–413, 2005, in Chinese.
- [7] M. C. He, H. P. Xie, and S. P. Peng, "Research on rock mass mechanics in deep mining," *Chinese Journal of Rock Mechanics and Engineering*, vol. 16, pp. 2803–2813, 2005, in Chinese.
- [8] X. T. Zhang and S. H. Chen, "Study on blast fragmentation for jointed and fractured rock mass considering collision," *Chinese Journal of Rock Mechanics and Engineering*, vol. 21, no. 8, pp. 1141–1146, 2002, in Chinese.
- [9] Z. Huang, X. Xu, H. Zhu, and M. Zhou, "An efficient group recommendation model with multiattention-based neural

- networks,” *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2020.
- [10] X. T. Zhang and Y. G. Ji, “Comparative analysis on blasting effect of different cutting blasting for the sandstone tunnel drivage in iron mine,” *Journal of the Balkan Tribological Association*, vol. 22, no. 1A, pp. 1014–1029, 2016.
- [11] Y. Sun, C. Xu, G. F. Li et al., “Intelligent human computer interaction based on non redundant EMG signal,” *Alexandria Engineering Journal*, 2020.
- [12] X. T. Zhang, Z. X. Li, and Q. Ma, “Study on the correlation between shpc pore structure and air permeability,” *Tehnicki Vjesnik-Technical Gazette*, vol. 24, no. 5, pp. 1425–1430, 2017.
- [13] X. Zhang, X. Zhou, H. Zhou, K. Gao, and Z. Wang, “Studies on forecasting of carbonation depth of slag high performance concrete considering gas permeability,” *Applied Clay Science*, vol. 79, pp. 36–40, 2013.
- [14] H. Li, X. Zhang, D. Li, L. Wu, W. Gao, and H. Zhou, “Numerical simulation of the effect of empty hole between adjacent blast holes in the perforation process of blasting,” *Journal of Intelligent & Fuzzy Systems*, vol. 37, no. 3, pp. 3137–3148, 2019.
- [15] Y. W. Lv and C. P. Sun, “Experimental study on damage evolution and crack propagation characteristics of sandstone under combined stress state,” *Journal of Shandong University of Science and Technology (Natural Science)*, vol. 39, no. 1, pp. 37–45, 2020, in Chinese.
- [16] Y. Sun, J. Xu, H. Qiang, and G. Lin, “Adaptive neural-fuzzy robust position control scheme for Maglev train systems with experimental verification,” *IEEE Transactions on Industrial Electronics*, vol. 66, no. 11, pp. 8589–8599, 2019.
- [17] K. Cui and X. Jing, “Research on prediction model of geotechnical parameters based on BP neural network,” *Neural Computing and Applications*, vol. 31, no. 12, 2019.
- [18] H. Sun, C. Gao, Z. Zhang, X. Liao, X. Wang, and J. Yang, “High-resolution anisotropic prestack Kirchhoff dynamic focused beam migration,” *IEEE Sensors Journal*, p. 1, 2019.
- [19] Z. Sun, S. Z. Xie, and X. T. Zhang, “Vibration reduction effect of damping holes in blasting excavation of tunnels with small spacing,” *Journal of Shandong University of Science and Technology (Natural Science)*, vol. 38, no. 1, pp. 25–31+39, 2019, in Chinese.
- [20] C. R. Li, L. J. Kang, and Q. X. Qi, “Probe into relationship between zonal fracturing and rock burst in deep tunnel,” *Journal of China Coal Society*, vol. 35, no. 2, pp. 185–189, 2016, in Chinese.
- [21] Y. N. He, L. J. Han, and P. Shao, “Some problems of rock mechanics for roadways stability in depth,” *Journal of China University of Mining & Technology*, vol. 3, pp. 288–295, 2006, in Chinese.
- [22] L. Hu, Y. J. Wang, and G. F. Ren, “Experimental study on the best pattern of blasthole of shale,” *Blasting*, vol. 32, no. 4, pp. 44–48, 2015, in Chinese.
- [23] Y. S. Lou and Y. Q. Jin, *Rock Mechanics and Petroleum Engineering*, Petroleum Industry Press, Beijing, China, 2006, in Chinese.
- [24] X. Huang, H. N. Ruan, and X. W. Wang, “Stability analysis and control of deep soft rock roadway under different lateral pressure coefficients,” *China Coal*, vol. 44, no. 7, pp. 59–63+87, 2018, in Chinese.
- [25] L. X. Xie, W. B. Lu, Q. B. Zhang, Q. H. Jiang, M. Chen, and J. Zhao, “Analysis of damage mechanisms and optimization of cut blasting design under high in-situ stresses,” *Tunnelling and Underground Space Technology*, vol. 66, pp. 19–33, 2017, in Chinese.
- [26] W. Wei, X. Xia, and W. Marcin, “Multi-sink distributed power control algorithm for Cyber-physical-systems in coal mine tunnels,” *Computer Networks*, vol. 161, pp. 201–219, 2019.

## Research Article

# A Bayesian Network under Strict Chain Model for Computing Flow Risks in Smart City

Zengfanxiang Wei <sup>1</sup>, Lei Zhang <sup>1</sup>, Qi Yue <sup>2</sup>, and Muchen Zhong <sup>3</sup>

<sup>1</sup>School of Economics and Management, Beijing Jiaotong University, Beijing 100044, China

<sup>2</sup>School of Information Management, Jiangxi University of Finance and Economics, Nanchang 330013, China

<sup>3</sup>Department of Cyber Security, Lancaster University, Lancaster LA2 0PF, Lancaster, UK

Correspondence should be addressed to Lei Zhang; 19125341@bjtu.edu.cn

Received 5 April 2020; Accepted 19 May 2020; Published 8 June 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Zengfanxiang Wei et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Risk management is a key factor for smart city running. There are many risk events in a strict process like transportation management of a smart city or a medical surgery in a smart hospital, and every step may lead to one kind of risk or more. In view of the fact that the occurrence of the flow risks follows the sequence formed by each process step, this paper presents a Bayesian network under strict chain (BN\_SC) to model this situation. In this model, the probabilistic reasoning formula is given according to the sequence of process steps, and the probabilities given by the model can do risk factor analysis to support the system to find an effective way to improve the process like machine manufacturing or a medical surgery. Finally, an example is analyzed based on the information given by doctors according to the situation of LC in their hospital located in Sichuan Province of China, which shows the effectiveness and rationality of the proposed BN\_SC model.

## 1. Introduction

Machine learning has applications in various fields for smart city running, such as machinery manufacturing, especially in probability evaluation [1–3]. Bayesian learning as a machine learning method can estimate the probabilities of risk occurrence like medical risk events in the medical process [4–7]. One or more of which will lead to the occurrence of other risk events, for example, in laparoscopic cholecystectomy, trocar puncture may cause intestinal injury, which often leads to infection.

In this kind of situation, the correlations of events can be quantified in a Bayesian network [8, 9]. However, the occurrence of risk events connected with the medical process, for example, LC must follow the surgical steps that construct a strict chain for the Bayesian net to follow. Existing quantitative research rarely studies on models for this special feature. According to this situation, we present a Bayesian network under strict chain (BN\_SC) model for the first time. In this model, the probabilistic reasoning formula according to the sequence of surgical steps processes the logical

inference for the probability of each risk subjected to the strict chain constructed by the surgical steps of LC. BN\_SC model can calculate the probabilities of risk flow more effectively than the traditional Bayesian network.

The remainder of the paper is organized as follows. Section 2 will review the research about the Bayesian network and we will illustrate the process of LC with a directed acyclic graph in Section 3. In Section 4, the BN\_SC model will be proposed, including the logical inference of the risks events and the formula to compute the probabilities of risks in LC. Finally, an example is analyzed based on the information given by doctors according to the situation of laparoscopic cholecystectomy in their hospital, which shows the effectiveness and rationality of the BN\_SC in Section 5.

## 2. Related Works

Previous studies have comprehensively expounded the concept and related properties of Bayesian networks. Tuyls and Maes pointed out that a Bayesian network is a directed acyclic graph, which consists of a set of random variables.

The meaning from node  $X$  to node  $Y$  is that node  $X$  has a direct influence on node  $Y$  [10]. The main feature of the Bayesian network is that it shows the interdependent relationship between variables. Koller and Friedman et al., Meinhshausen and Btihlmann, and Wainwright and Jordan pointed out that the graph connected by undirected edges between variables is an undirected graph known as Markov network [11–13]. These graphs produce undirected graph models, which are more suitable for analyzing the similarity and related behavior between variables. Ghahramani, Heckerman et al., and Neapolitan pointed out that a graph connected by a directed edge of variables is a “directed graph,” and a directed graph without a cycle is a “directed acyclic graph” (DAG) [14–16]. Directed acyclic graphs are usually based on the concept of family order. For example, in  $A \rightarrow B \rightarrow C$ ,  $A$  is the parent node of  $B$ ,  $C$  is the child node of  $B$ ,  $A$  and  $B$  are the ancestors of  $C$ , and  $B$  and  $C$  are the descendants of  $A$ . This ranking clearly expresses causality, and these figures produce a class of models called Bayesian networks. Andersson et al., Gillispie and Perlman, and Pearl have pointed out that some directed acyclic graphs have two-way edges, which are suitable for situations where some variables interact with each other [17–19]. They represent a class of Markov equivalent directed acyclic graphs.

Bayesian network is widely used in risk evaluation research. Ronald et al. show that chains of events can lead to increased or critical emissions of volatile organic compounds by assessing risks based on the Bayesian network [20]. Bhattacharjee applied a Bayesian model that can handle death data analysis without stratifying in the presence of competing risks [21]. Sanchez et al. develop a general framework and a method to estimate the impact of project management maturity on project performance by using Bayesian networks to formalize project management experts’ knowledge [22]. Masmoudi et al. used a discrete Bayesian network with a latent variable to model the payment default of loan subscribers [23]. Punyamurthula and Badurdeen represented operational risks of production line and their causal relationships based on Bayesian Belief Networks [24].

### 3. Process of LC

Laparoscopic surgery is a new method of minimally invasive surgery, which is the inevitable trend of the future development of surgical methods. Cholecystectomy is one of the most common abdominal operations, which is usually performed in laparoscopic surgery in developed countries. For example, 90% of cholecystectomy in the United States is performed by laparoscopy. Laparoscopic cholecystectomy is regarded as the “gold standard” for surgical treatment of cholelithiasis [25, 26].

Laparoscopic cholecystectomy is divided into five steps, which include trocar puncture (into the abdomen), separation of adhesion, anatomy bile triangle, dissection of gallbladder bed, and removal and suture of gallbladder. In the surgery, intestinal injury may be caused by trocar puncture, bile duct injury may be caused by separation of adhesion, anatomy of bile triangle, and dissection of

gallbladder, and vascular injury may be caused by removal and suture of gallbladder. Moreover, the three kinds of injuries may lead to infection and massive hemorrhage, which can eventually lead to death. In order to describe the risk probability in LC surgery more succinctly, Table 1 shows the mathematical symbols in this paper.

According to analysis, we can get the directed acyclic graph of LC operation as shown in Figure 1.

Figure 1 describes the specific process how the LC works. We can build a special model by following the logic derived from the strict chain. As shown in the figure, the risk of  $R_2$  happened in the process of Step 2. Moreover, the occurrence of  $R_1$  means the failure of Step 1, which means the surgery cannot be performed and  $R_2$  cannot occur. The five surgery steps formed a chain above the Bayesian network of risks in the surgery. The occurrence of each risk needs to follow the order of the steps in the chain strictly. This changes the rule of the traditional inference of Bayesian network. To solve the problem, we present a new model in Section 4.

### 4. Bayesian Network under Strict Chain (BN\_SC)

According to the analysis in Section 2, we build a Bayesian network to model the LC process. Different from the traditional Bayesian network, in many medical procedures, the risk Bayesian network is developed under the surgical process such as the Bayesian network of LC surgery in Section 2, and the surgical process is characterized by the fact that the next step can only be taken after the previous step surgical procedure is successful. So those medical steps construct a strict process chain naturally such as  $S_1$  to  $S_5$  of LC, and more importantly, the Bayesian network of surgery risks should follow the steps of the chain model. According to this characteristic, this paper proposes Bayesian network under strict chain model. Under BN\_SC model, if  $j = i + 1$ , then surgical step  $j$  can only begin after surgical Step  $i$  completes successfully. The formula for calculating the probability of the risk occurrence of the event  $j$  is as follows:

$$P(S_j) = \prod_{i=1}^{j-1} P(S_i)P(S_i). \quad (1)$$

Here,  $P(D_i)$  means the probability of the surgical Step  $i$  completed successfully. This paper assumed that only when the surgical Step  $i$  causes the risk  $i$  will lead to the unsuccessful surgery, and some other unexpected factors such as power outages are not under consideration. So,  $P(D_i)$  and  $P(R_j)$  form complementary relationship, that is,  $P(R_j) = 1 - P(D_i)$ . In laparoscopic surgery, the patient is bound to have the first surgical step after entering the procedure, so the first surgical step is bound to occur here, that is,  $P(S_1) = 1$ .

Therefore, the probability of intestinal injury risk  $P(R_1)$  caused by the first surgical procedure can be represented as follows:

$$P(R_1) = P(R_1 | S_1)P(S_1) = P(R_1 | S_1). \quad (2)$$

TABLE 1: List of symbol notations and description.

| Symbols         | Descriptions   |
|-----------------|--|
| $S_1$           | Trocar puncture (into the abdomen) happened              |
| $S_2$           | Separation of adhesion happened                          |
| $S_3$           | Anatomy of bile triangle happened                        |
| $S_4$           | Dissection of gallbladder bed happened                   |
| $S_5$           | Removal and suture of gallbladder happened               |
| $D_1$           | Trocar puncture (into the abdomen) has done successfully |
| $D_2$           | Separation of adhesion has done successfully             |
| $D_3$           | Anatomy of bile triangle has done successfully           |
| $D_4$           | Dissection of gallbladder bed has done successfully      |
| $D_5$           | Removal and suture of gallbladder has done successfully  |
| $R_1$           | Intestinal injury  |
| $R_2$           | Bile duct injury   |
| $R_3$           | Vascular injury  |
| $R_4$           | Infection  |
| $R_5$           | Massive hemorrhage                                       |
| $R_6$           | Death  |
| $W_1, W_2, W_3$ | Contribution ratio of $R_1, R_2, R_3$ to $R_4$           |
| $W_4, W_5, W_6$ | Contribution ratio of $R_1, R_2, R_3$ to $R_5$           |

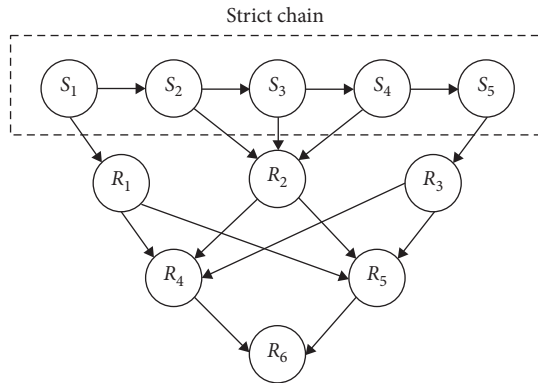


FIGURE 1: Bayesian network under a strict chain.

The probability of bile duct injury risk  $P(R_2)$  caused by multiple surgical procedures  $S_2$ ,  $S_3$ , and  $S_4$  can be represented as follows:

$$\begin{aligned}
 P(R_2) &= P(R_2 | S_2)P(S_2) + P(R_2 | S_3)P(S_2 | S_3)P(S_3) + \\
 &\quad P(R_2 | S_4)P(S_4 | S_3)P(S_3 | S_2)P(S_2) \\
 &= (1 - P(D_2))P(D_1) + (1 - P(D_3))P(D_2)P(D_1) + \\
 &\quad (1 - P(D_4))P(D_3)P(D_2)P(D_1) \\
 &= P(D_1)(1 - P(D_2)P(D_3)P(D_4)).
 \end{aligned} \tag{3}$$

Moreover, the probability of incision tear risk  $P(R_3)$  can be represented as follows:

$$\begin{aligned}
 P(R_3) &= P(R_3 | S_5)P(S_5) \\
 &= P(R_3 | S_5)P(D_4)P(D_3)P(D_2)P(D_1).
 \end{aligned} \tag{4}$$

According to the DAG of LC, three risks can lead to the risk of infection  $R_4$  and the risk of massive hemorrhage  $R_5$ . However, this conditional probability cannot be calculated by statistics. It can only be quantitatively judged based on the experience accumulated by doctors in completed cases. In order to get this judgement value, two sets of comparison matrices are given by doctors who are allowed to score 1–9. The numbers 3, 5, 7, and 9 correspond to the verbal judgments “moderately more dominant,” “strongly more dominant,” “very strongly more dominant,” and “extremely more dominant” (with 2, 4, 6, and 8 for compromise between the previous values). We are permitted to interpolate values between the integers, if desired, or use numbers from an actual ratio scale of measurement. This scale is mathematically derived from stimulus-response theory and has been extended through the use of structuring and decomposition to assume arbitrarily large values as necessary. Firstly, for risk  $R_4$ , we use pairwise comparison to research which risk  $R_1, R_2, R_3$  is more likely to cause  $R_4$ . After entering scores given by doctors into matrix, the contribution ratio of three kinds of risk  $R_1, R_2, R_3$  to  $R_4$  is obtained by calculating the eigenvector, and the contribution ratio of  $R_1, R_2, R_3$  can be represented as  $w_1, w_2, w_3$ :

$$\begin{array}{c} R_4 \\ \hline R_1 \ R_2 \ R_3 \end{array} \begin{pmatrix} R_{ij} \end{pmatrix} \rightarrow \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}. \tag{5}$$

By the same token, the contribution ratio of  $R_1, R_2, R_3$  to  $R_5$  can be represented as  $w_4, w_5, w_6$ :

$$\begin{array}{c} R_5 \\ \hline R_1 \ R_2 \ R_3 \end{array} \begin{pmatrix} R_{ij} \end{pmatrix} \rightarrow \begin{pmatrix} w_4 \\ w_5 \\ w_6 \end{pmatrix}. \tag{6}$$

In addition, by comparing the risk of massive bleeding and the risk of infection leading to death, the weight of risk  $R_4, R_5$  can be obtained, which can be represented by  $C_0, C_1$ :

$$\begin{array}{c} R_6 \\ \hline R_4 \ R_5 \end{array} \begin{pmatrix} R_{ij} \end{pmatrix} \rightarrow \begin{pmatrix} C_0 \\ C_1 \end{pmatrix}. \tag{7}$$

Then, the final weight  $C_2, C_3, C_4$  of  $R_1, R_2, R_3$  can be obtained by the following formula:

$$C_2 = C_0 w_1 + C_1 w_4, \tag{8}$$

$$C_3 = C_0 w_2 + C_1 w_5, \tag{9}$$

$$C_4 = C_0 w_3 + C_1 w_6. \tag{10}$$

Based on the above formula, we can use  $P(D_1), P(D_2), P(D_3), P(D_4), P(D_5)$  to express the probability of death. According to the following formula, we

analyze the relations between each surgery step and death risk:

$$\begin{aligned}
P(R_6) &= C_0P(R_4) + C_1P(R_5) = C_2P(R_1) \\
&\quad + C_3P(R_2) + C_4P(R_3) \\
&= C_2(1 - P(D_1)) + C_3P(D_1)(1 - P(D_2)P(D_3)P(D_4)) \\
&\quad + C_4(1 - P(D_5))P(D_4)P(D_3)P(D_2)P(D_1) \\
&= C_2 + (C_3 - C_2)P(D_1) + (C_4 - C_3)P(D_1)P \\
&\quad \cdot (D_2)P(D_3)P(D_4) \\
&\quad - C_4P(D_1)P(D_2)P(D_3)P(D_4)P(D_5).
\end{aligned} \tag{11}$$

## 5. Application: Validating Our BN\_SC Model

To further illustrate the calculation process of proposed method, we implement the BN\_SC model to a hospital in Sichuan Province of China. According to the existing laparoscopic cholecystectomy records of surgical cases, the hospital counted 1166 cases in the past year. The statistical results demonstrate the success ratios of trocar puncture, separation of adhesion, anatomy of bile triangle, dissection of gallbladder bed, and removal and suture of gallbladder were 0.990, 0.970, 0.873, 0.830, and 0.821. The probability of intestinal injury is 0.010 according to formula (2), and the probability of bile duct injury and vascular injury is 0.294 and 0.125 according to formula (3).

In addition, information about death, infection, and massive bleeding involves the privacy of patients, so doctors can only judge the relationship between them based on experience instead of providing specific values. According to the method given in Section 3, we asked the attending physician of laparoscopic cholecystectomy operation to give the subjective evaluation as follows:

Three surgical risk comparison matrices for infection are described as follows:

$$\begin{array}{c|ccc}
R_4 & R_1 & R_2 & R_3 \\
\hline
R_1 & 1 & 1/6 & 1/4 \\
R_2 & 6 & 1 & 3 \\
R_3 & 4 & 1/3 & 1
\end{array} \tag{12}$$

By finding the eigenvector of the matrix, the weight is obtained:

$$\begin{pmatrix} w_1 = 0.09 \\ w_2 = 0.64 \\ w_3 = 0.27 \end{pmatrix}. \tag{13}$$

Three surgical risk comparison matrices for massive hemorrhage are described as follows:

$$\begin{array}{c|ccc}
R_5 & R_1 & R_2 & R_3 \\
\hline
R_1 & 1 & 3 & 1/5 \\
R_2 & 1/3 & 1 & 1/7 \\
R_3 & 5 & 7 & 1
\end{array} \tag{14}$$

By finding the eigenvector of the matrix, the weight is obtained:

$$\begin{pmatrix} w_4 = 0.19 \\ w_5 = 0.08 \\ w_6 = 0.73 \end{pmatrix}. \tag{15}$$

The surgical risk comparison matrices for death are described as follows:

$$\begin{array}{c|cc}
R_6 & R_4 & R_5 \\
\hline
R_4 & 1 & 1/2 \\
R_5 & 2 & 1
\end{array} \tag{16}$$

By finding the eigenvector of the matrix, the weight is obtained:

$$\begin{aligned}
C_0 &= 0.333, \\
C_1 &= 0.666.
\end{aligned} \tag{17}$$

We can obtain  $C_2, C_3, C_4$  by formulas (8)–(10):

$$\begin{aligned}
C_2 &= 0.157, \\
C_3 &= 0.267, \\
C_4 &= 0.576.
\end{aligned} \tag{18}$$

According to the analysis, the death risk of laparoscopic cholecystectomy surgery in a hospital in Sichuan can be obtained as follows:

$$\begin{aligned}
P(R_6) &= C_0P(R_4) + C_1P(R_5) = C_2P(R_1) + C_3P(R_2) + C_4P(R_3) \\
&= C_2 + (C_3 - C_2)P(D_1) + (C_4 - C_3)P(D_1)P \\
&\quad \cdot (D_2)P(D_3)P(D_4) \\
&\quad - C_4P(D_1)P(D_2)P(D_3)P(D_4)P(D_5) \\
&= 0.157 \times 0.010 + 0.267 \times 0.294 + 0.576 \times 0.125 \\
&= 0.152.
\end{aligned} \tag{19}$$

The BN\_SC model is more effective than the traditional Bayesian model in reducing the computations of the particular conditional probabilities, because some of the conditional probabilities can be replaced by the probabilities of ex-step.

Sensitivity analysis for the quantitative relationships between the success rates of each laparoscopic cholecystectomy surgery step and the probability of death are as follows from Figure 2 to Figure 11:

$$\begin{aligned}
P(R_6) &= 0.157 - 0.005P(D_1), \\
P(R_6) &= 0.266 - 0.117P(D_2), \\
P(R_6) &= 0.266 - 0.130P(D_3), \\
P(R_6) &= 0.266 - 0.137P(D_4), \\
P(R_6) &= 0.215 - 0.077P(D_5).
\end{aligned} \tag{20}$$

Through the above experimental results, it can be observed that the most effective way to reduce the risk of death is to improve the success rate of dissection of gallbladder, and the second is to improve the success rate of anatomical



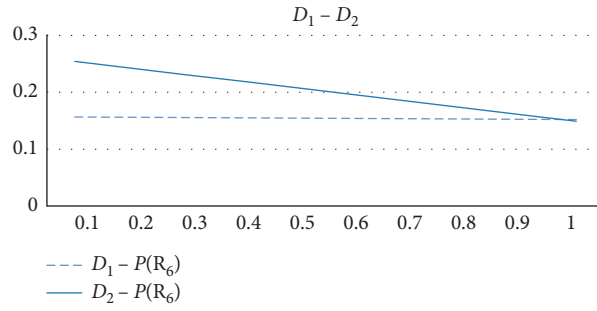


FIGURE 2: Sensitivity analysis of  $D_1 - D_2$ .

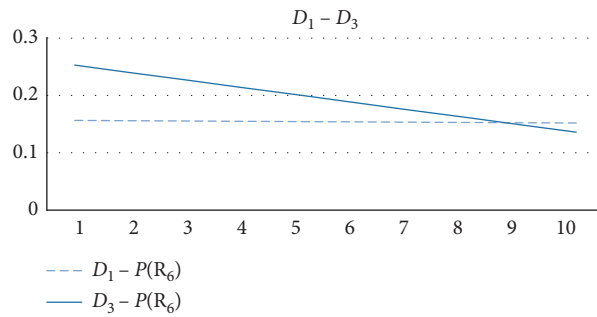


FIGURE 3: Sensitivity analysis of  $D_1 - D_3$ .

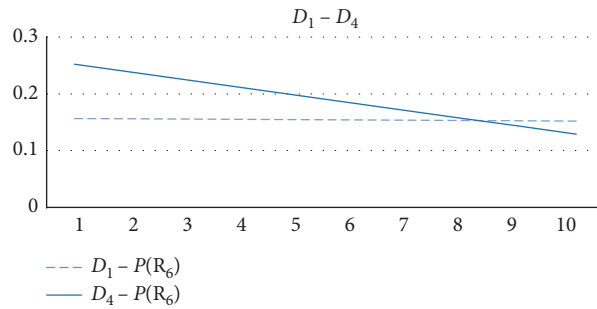


FIGURE 4: Sensitivity analysis of  $D_1 - D_4$ .

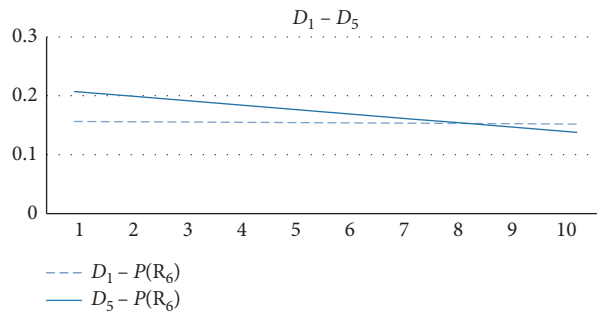
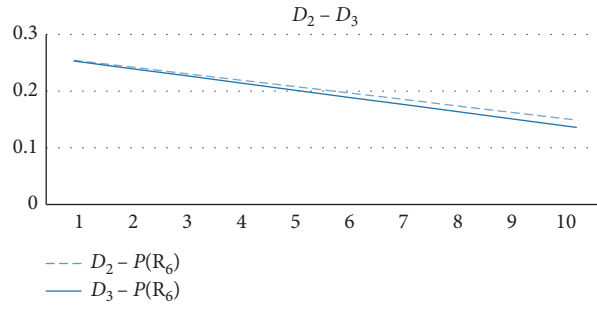
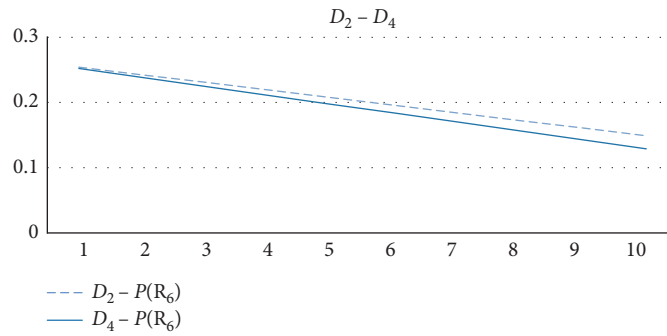
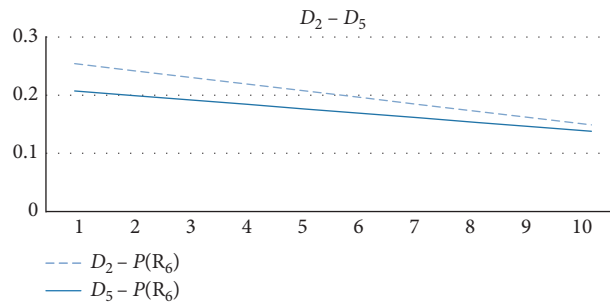
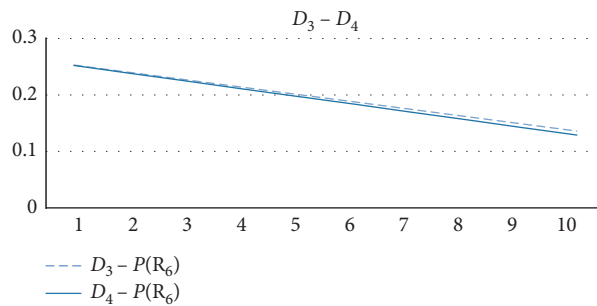
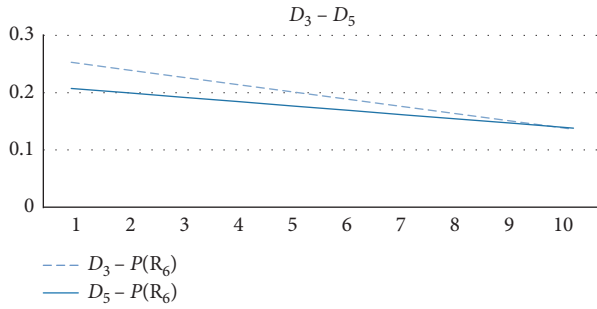
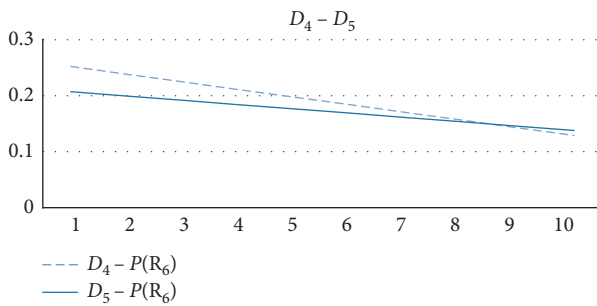


FIGURE 5: Sensitivity analysis of  $D_1 - D_5$ .

FIGURE 6: Sensitivity analysis of  $D_2 - D_3$ .FIGURE 7: Sensitivity analysis of  $D_2 - D_4$ .FIGURE 8: Sensitivity analysis of  $D_2 - D_5$ .FIGURE 9: Sensitivity analysis of  $D_3 - D_4$ .

FIGURE 10: Sensitivity analysis of  $D_3 - D_5$ .FIGURE 11: Sensitivity analysis of  $D_4 - D_5$ .

bile triangle. In addition, the least efficient way to reduce the risk of death is to improve the success rate of trocar puncture surgery. These conclusions have certain guiding significance for the improvement of laparoscopic surgery of a hospital in Sichuan.

## 6. Conclusion

As a crucial factor of running smart city, risk flow management need to calculate accurately. Our BN\_SC model modeled a special situation for risks flow in Bayesian net, which logical inference has to follow a strict chain. This is to describe surge process especially LC in this paper. The BN\_SC computed the probabilities of risks in LC by Bayesian inference follow the particular process. Our model is a special Bayesian model for the specific system, limiting to the structure of the process, but can reduce computations in the suitable situations by the formula we presented.

The application of the BN\_SC has proved its rationality and practicality. The method also has a way to combine the judgments and data of several hospitals to representative group judgments to deal with the weights of risks. Our model can be applied in other probabilities inference problems that have a strict chain to follow; for example, stock prices follow a timeline or story events follow a plotline.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was partly supported by the National Natural Science Foundation of China (Grant no. 71861015), Distinguished Young Scholar Talent of Jiangxi Province (Grant no. 20192BCBL23008), the Humanities and Social Science Foundation of the Ministry of Education of China (Grant no. 18YJA630047), and the Humanities and Social Science Foundation of Universities of Jiangxi Province (Grant no. GL18122).

## References

- [1] A. Gelman and C. P. Robert, ““Not only defended but also applied”: the perceived absurdity of bayesian inference,” *The American Statistician*, vol. 67, no. 1, pp. 1–5, 2013.
- [2] J. Yang, N. Xiong, A. V. Vasilakos et al., “A fingerprint recognition scheme based on assembling invariant moments for cloud computing communications,” *IEEE Systems Journal*, vol. 5, no. 4, pp. 574–583, 2011.
- [3] H. Zhenhua, X. Xin, Z. Honghao, and Z. MengChu, “An efficient group recommendation model with multiattention-based neural networks,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, pp. 1–14, 2020.
- [4] R. Jiang, M. Shi, and W. Zhou, “A privacy security risk analysis method for medical big data in urban computing,” *IEEE Access*, vol. 7, no. 12, pp. 143841–143854, 2019.
- [5] J. Hu, Y. Sun, G. Li, G. Jiang, and B. Tao, “Probability analysis for grasp planning facing the field of medical robotics,” *Measurement*, vol. 141, pp. 227–234, 2019.
- [6] Q. Yue, L. Zhang, B. Yu, L. J. Zhang, and J. Zhang, “Two-sided matching for triangular intuitionistic fuzzy numbers in smart environmental protection,” *IEEE Access*, vol. 7, pp. 42426–42435, 2019.
- [7] Q. Yue and L. Zhang, “Two-sided matching for hesitant fuzzy numbers in smart intelligent technique transfer,” *Mechanical Systems and Signal Processing*, vol. 139, p. 106643, 2020.
- [8] J. Sampaio, E. J. Drinkwater, and N. M. Leite, “Effects of season period, team quality, and playing time on basketball players’ game-related statistics,” *European Journal of Sport Science*, vol. 10, no. 2, pp. 141–149, 2010.
- [9] A. C. Constantinou, N. E. Fenton, and M. Neil, “Profiting from an inefficient association football gambling market: prediction, risk and uncertainty using bayesian networks,” *Knowledge-Based Systems*, vol. 50, no. 2, pp. 60–86, 2013.
- [10] K. Tuyls, S. Maes, and B. Vanschoenwinkel, “Machine learning techniques for fraud detection,” Master thesis, VUB, Brussels, Belgium, 2000.
- [11] D. Koller, N. Friedman, and F. Bach, *Probabilistic Graphical Models: Principles and Techniques*, MIT press, Cambridge, MA, USA, 2009.
- [12] N. Meinshausen and P. Bühlmann, “High-dimensional graphs and variable selection with the Lasso,” *The Annals of Statistics*, vol. 34, no. 3, pp. 1436–1462, 2006.
- [13] M. J. Wainwright and M. I. Jordan, “Graphical models, exponential families, and variational inference,” *Foundations and Trends® in Machine Learning*, vol. 1, no. 1–2, pp. 1–305, 2008.
- [14] Z. Ghahramani, “Learning dynamic bayesian networks,” in *International School on Neural Networks, Initiated by IIASS and EMFCSC*, pp. 168–197, Springer, Berlin, Germany, 1997.
- [15] D. Heckerman, D. Geiger, and D. M. Chickering, “Learning bayesian networks: the combination of knowledge and

- statistical data,” *Machine Learning*, vol. 20, no. 3, pp. 197–243, 1995.
- [16] R. E. Neapolitan, *Learning Bayesian networks*, Pearson Prentice Hall, Upper Saddle River, NJ, USA, 2004.
- [17] S. A. Andersson, D. Madigan, and M. D. Perlman, “A characterization of markov equivalence classes for acyclic digraphs,” *The Annals of Statistics*, vol. 25, no. 2, pp. 505–541, 1997.
- [18] S. B. Gillispie and M. D. Perlman, “Enumerating markov equivalence classes of acyclic digraph dels,” in *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann Publishers Inc, Seattle, WA, USA, pp. 171–177, August 2001.
- [19] J. Pearl, *Causality: Models, Reasoning and Inference*, Cambridge University Press, Cambridge, UK, 2000.
- [20] R. Zinke, J. Melnychuk, F. Köhler, and U. Krause, “Quantitative risk assessment of emissions from external floating roof tanks during normal operation and in case of damages using Bayesian Networks,” *Reliability Engineering & System Safety*, vol. 197, p. 106826, 2020.
- [21] A. Bhattacharjee, “Bayesian competing risks analysis without data stratification,” *Clinical Epidemiology and Global Health*, vol. 8, no. 1, pp. 265–270, 2020.
- [22] F. Sanchez, E. Bonjour, J.-P. Micaelli, and D. Monticolo, “An approach based on bayesian network for improving project management maturity: an application to reduce cost overrun risks in engineering projects,” *Computers in Industry*, vol. 119, p. 103227, 2020.
- [23] K. Masmoudi, L. Abid, and A. Masmoudi, “Credit risk modeling using Bayesian network with a latent variable,” *Expert Systems with Applications*, vol. 127, no. 8, pp. 157–166, 2019.
- [24] S. Punyamurthula and F. Badurdeen, “Assessing production line risk using bayesian Belief networks and system dynamics,” *Procedia Manufacturing*, vol. 26, no. 8, pp. 76–86, 2018.
- [25] J. H. Lee, S. Kim, I. Lee et al., “A risk prediction model for medical treatment failure in tubal pregnancy,” *European Journal of Obstetrics & Gynecology and Reproductive Biology*, vol. 225, pp. 148–154, 2018.
- [26] S. Stonelake, P. Thomson, and N. Suggett, “Outcomes following emergency laparotomy: a comparison of predicted possum morbidity with the Clavien-Dindo classification of surgical complications,” *International Journal of Surgery*, vol. 11, no. 8, p. 717, 2013.

## Research Article

# Research on Sentiment Tendency and Evolution of Public Opinions in Social Networks of Smart City

Yanni Liu,<sup>1</sup> Dongsheng Liu ,<sup>2</sup> and Yuwei Chen<sup>3</sup>

<sup>1</sup>School of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou 310018, China

<sup>2</sup>School of Computer and Information Engineering, Zhejiang Gongshang University, Hangzhou 310018, China

<sup>3</sup>Beijing Yunzhenxin Technology Co., Ltd., Hangzhou 310012, China

Correspondence should be addressed to Dongsheng Liu; [lds1118@zjgsu.edu.cn](mailto:lds1118@zjgsu.edu.cn)

Received 29 March 2020; Revised 27 April 2020; Accepted 5 May 2020; Published 4 June 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Yanni Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of mobile Internet, the social network has become an important platform for users to receive, release, and disseminate information. In order to get more valuable information and implement effective supervision on public opinions, it is necessary to study the public opinions, sentiment tendency, and the evolution of the hot events in social networks of a smart city. In view of social networks' characteristics such as short text, rich topics, diverse sentiments, and timeliness, this paper conducts text modeling with words co-occurrence based on the topic model. Besides, the sentiment computing and the time factor are incorporated to construct the dynamic topic-sentiment mixture model (TSTS). Then, four hot events were randomly selected from the microblog as datasets to evaluate the TSTS model in terms of topic feature extraction, sentiment analysis, and time change. The results show that the TSTS model is better than the traditional models in topic extraction and sentiment analysis. Meanwhile, by fitting the time curve of hot events, the change rules of comments in the social network is obtained.

## 1. Introduction

With the wide application of the Internet technology, the Internet has gradually transformed to the dynamic platform for information sharing and interactive communication. The 43<sup>rd</sup> statistical report indicated that China had 854 million Internet users, and 99.1 percent of them access the Internet via mobile phones [1]. Social networks of a smart city have become the mainstream platform for information exchange and opinion expression. The users are not only the receivers of information, but also the creators to publish text comments in social networks. The hot events of public opinion refer that personal opinions are released on upcoming or already happened events by online communication tools and network platforms [2]. The spread of public opinion will snowball and expand by social networks, and emergent events may develop in an uncontrollable direction. Chain events caused by inadequate supervision on social networks can bring about the bad influence, and the frequency and

harmfulness have shown an obvious rising trend in recent years [3].

Previous research has been studied from the qualitative aspects, such as the evolution mechanism of public opinion, information element classification, and influence judgment. However, the above research cannot meet the needs of online public opinion supervision, and the monitoring and management of hot events in social networks of a smart city need to implement quantitative judgment. For public opinion monitoring and management, Steyvers and Griffiths [4] proposed a topic model for public opinion detection in the social network. Yeh et al. [5] proposed a conceptually dynamic latent Dirichlet allocation (CD-LDA) model for topic content detection and tracking. The studies on probabilistic topic models for extracting hot topics from long texts have achieved good results [6], but these models are not suited to extract hot topics from short texts, such as Twitter and Facebook [7]. Kim et al. [8] introduced the sentiment scoring based on topics through the n-gram LDA

topic modeling technology and investigated the topic reports and sentiment dynamics in news about Ebola virus. Subeno et al. [9] proposed a collapsed Gibbs sampling method based on the latent Dirichlet allocation (LDA) model widely used on Spark. Park et al. [10] used partially collapsed Gibbs sampling for latent Dirichlet allocation and proposed a reasoning LDA method, which effectively obtained unbiased estimation under flexible modeling of heterogeneous text corpus by partially collapsed and Dirichlet mixed processing.

However, there are still some problems in the detection of public opinion for events in social networks of a smart city. Firstly, the detection and analysis of public opinions for hot events in social networks mostly remain in the qualitative analysis or empirical research, lacking quantitative research. Secondly, there is a lack of public opinion analysis method combining with the characteristics of the microblog in social networks. Thirdly, for the sentiment analysis of public opinion events, most research adopts the two-stage method. That is to detect the event firstly and then conduct sentiment analysis and judgment, which is likely to lead to the separation between the event and sentiment. Fourthly, the dissemination for public opinions of hot events is time sensitive, so it is necessary to involve the time factor in comment text analysis. Thus, this paper proposes a mixed model with dynamic topic-sentiment (TSTS) for short texts in the social network, which comprehensively incorporates the topics, sentiment, and time factor of events to detect the public opinion. By quantitative analysis of real experiment data, the model can not only show the quantitative evolution trend of public opinion, but also provide the propagation rule of sudden events.

The main contributions of this paper are reflected in two aspects. Firstly, the TSTS model is proposed by extending the topic model, which can not only extract both topic and sentiment polarity words, but also integrate the time factor to realize dynamic analysis of short texts. Secondly, this paper studies the detection and evolution analysis of Internet public opinion by the dynamic topic-sentiment model. The real datasets are used to conduct experimental analysis of the proposed model, which can reflect the evolution trend of public opinion diffusion.

## 2. Related Research

*2.1. Research on Relevant Topic Models.* The traditional opinion mining is to analyze the sentiment orientations based on the level of document and sentence. The traditional topic model was mainly used to compare the similarity among articles by comparing the number of repeated words in different articles. Blei et al. [11] proposed the latent Dirichlet allocation (LDA) topic model to mine the hidden semantics of the text. LDA is a three-layer Bayesian model involving the document, topic, and word. The document is composed of a mixed distribution of topics, and each topic follows a polynomial distribution. And, Dirichlet distribution is introduced as the prior information of the polynomial distribution. The schematic diagram of the LDA topic model is shown in Figure 1.

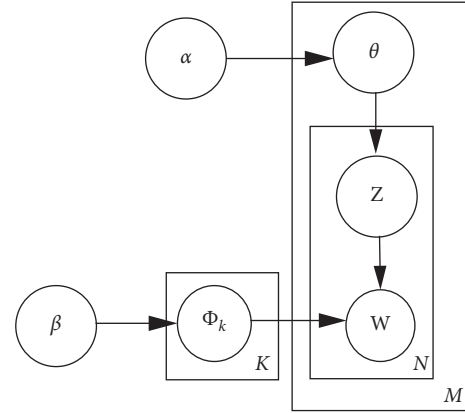


FIGURE 1: Schematic diagram of the LDA model.

*2.1.1. Research on Topic Model of Short Text.* For most topic models, the topic is the word that appears in the document and has some connection. In previous studies, the topic model of short texts was expanded by introducing relevant background or author information, which weakened the topic and produced meaningless word contributions. Similarly, if co-occurrence words are extended to the whole corpus in the experiment, the occurrence frequency of each word will be greatly increased, and the connection between words will be closer. Then, the modeling on documents will be easier. Based on the above hypotheses, Cheng et al. [12] proposed the biterm topic model (BTM), which is another way to explain the relationship between words, and text modeling of documents can be conducted based on the word co-occurrence model of the whole corpus. Rashid et al. [13] proposed the fuzzy topic modeling (FTM) for short texts from the fuzzy perspective to solve the sparsity problem. Based on BTM, Lu et al. [14] introduced the RIBS-Bigrams model by learning the usage relationship, which showed topics with two-letter groups. Zhu et al. [15] proposed a joint model based on the latent Dirichlet allocation (LDA) and BTM, which not only alleviates the sparsity of the BTM algorithm in processing short texts, but also preserves topic information of the document through the extended LDA.

*2.1.2. Research on Mixed Model Integrating Topic Sentiment.* To evaluate the sentiment tendency of documents, the joint sentiment topic (JST) model added the sentiment layer based on the LDA model, forming a four-layer Bayesian network [16]. In this structure, the sentiment polarity label is related to the document, and the word generation is also influenced by both topic and sentiment. In the traditional LDA model, the generation of the document and words is determined by the topic. But in the JST model, the word of the document is determined by the topic and the sentiment. Amplayo et al. [17] proposed the aspect and sentiment unification model (ASUM) with sentiment level. The difference between JST and ASUM is that words in a sentence come from different topics in the JST model, while all words of a sentence belong to one topic in the ASUM model.

*2.1.3. Research on Topic Model with Time Factor.* Yao et al. [18] revealed the semantic change process of words by correlating the time factor with Wikipedia text knowledge. In terms of event evolution, the associative topic model (ATM) is proposed [19], and the recognized cluster is represented as the word distribution of the cluster with the corresponding event. In addition, Topic Over Time (TOT) was proposed to integrate the time factor into the LDA model [20]. In the TOT model, word co-occurrence can affect the discovery of subject words, and time information can also affect the extraction of topic words. Unlike other models, each topic is subject to a continuous distribution of time and not rely on Markov models to discretize time in the TOT model. For each document generation, the mixed distribution of topics is determined by word co-occurrence and timestamp [21], which allows the TOT model to maintain independence in the time dimension and can predict the time for the document without any time information.

*2.2. Gibbs Sampling.* The derivation of the experimental model in the paper is a variant form of the Markov Chain, so the Markov Chain Monte Carlo (MCMC) method is used for sampling in the experiment. Gibbs sampling, as one of the MCMC methods, has been widely used in prior research. Gibbs sampling is used to obtain a set of observations that approximate a specified multidimensional probability distribution, such as the probability distribution of two random variables.

The Gibbs sampling method used for the latent Dirichlet allocation (LDA) model can significantly improve the speed of the real-text corpus [22]. Papanikolaou et al. [23] estimated latent Dirichlet allocation (LDA) parameters from Gibbs sampling by using all conditional distributions of potential variable assignments to effectively average multiple samples. Zhou et al. [24] proposed two kinds of Gibbs sampling inference methods, such as Sparse BTM and ESparse BTM, to achieve BTM by weighing space and time. Bhuyan [25] proposed the correlation random effect model based on potential variables and an algorithm to estimate correlation parameters based on Gibbs sampling.

### 3. Model Constructing

*3.1. Topic-Sentiment Mixture Model with Time Factor (TSTS).* Based on prior research, this paper mainly improves the topic model from three aspects. Firstly, the sparse matrix caused by short texts in the social network is solved. Secondly, the topic and sentiment distribution of the same word pair are controlled. Thirdly, the problem of text homogeneity is solved by incorporating the time factor into the topic model. Therefore, the TSTS model proposed in this paper is used to constrain the word pairs in the same document, which greatly reduces the complexity of time and space and makes up for the sparse matrix of short texts to some extent. Moreover, the sentiment layer is integrated into TSTS by extending the hypotheses of ASUM and restrains word pairs generated by constraining sentences to follow the same

topic-sentiment distribution. Finally, the TSTS model incorporating the time factor does not rely on the Markov model to discretize time, and each topic is subject to the continuous temporal distribution. For each document generation, the mixed distribution of topics is determined by the words co-occurrence and timestamps. TSTS model is shown in Figure 2.

The TSTS model simulates the generating process of online comments. Generally, the online comments from users can be regarded as a document, which is short, pithy, and highly emotional. The word co-occurrence from BTM is the most effective solution for the short-text topic model. In addition, the TSTS model with the time layer can continuously sample users' evaluation of hot events, as well as the dynamic changes of users' sentiment. Therefore, the hypotheses of the TSTS model are proposed as follows:

- (i) The probability distribution of the time factor is not directly equal to the joint distribution of the topic and sentiment
- (ii) The topic-sentiment distribution of each document is independent [26]
- (iii) Similar topics of different sentiment polarity are not automatically categorized [27]

Combined with the probability graph of Bayesian's network, the TSTS model proposed in the paper has four characteristics. First, a word pair is used to replace a single word to carry out the sampling model. Second, each timestamp is related by topic and sentiment. Third, the extraction of thematic characteristic and sentiment words is for the whole corpus. Fourth, in the derivation process of the TSTS model, it is not necessary to correspond between thematic feature and affective polarity words. That is because every topic and sentiment have the corresponding polynomial word pair distribution. In addition, the text modeling process of the TSTS model also follows the assumption that there is a connection between the sentimental polarity words of the topic features, which also changes with the time factor. So, the documents used to train the model must have a specific timestamp, such as the publishing time of the microblog.

*3.2. Generation of a Text in TSTS Model.* In the TSTS model, we assumed that a corpus is composed of several texts. For instance, a microblog is a text containing two dimensions of topic and sentiment. Considering the effectiveness of public opinions and related parameters of the microblog text, word distribution is determined by the topic, sentiment, and time. So, TSTS is an unsupervised topic-sentiment mixed model. The generation process of the document is as follows:

- (1) Extract a polynomial distribution  $\theta_d$  on a topic from the Dirichlet prior distribution  $\alpha$ , that is,  $\theta_d \sim \text{Dir}(\alpha)$
- (2) Extract a polynomial distribution  $\psi_{z,l}$  at some point from the Dirichlet prior distribution  $\mu$ , that is,  $\psi_{z,l} \sim \text{Dir}(\mu)$

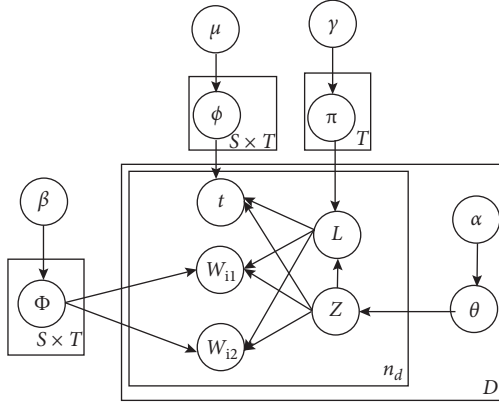


FIGURE 2: TSTS model.

- (3) Extract a polynomial distribution  $\pi_z$  in a sentiment from the Dirichlet prior distribution  $\gamma$ , that is,  $\pi_z \sim \text{Dir}(\gamma)$
- (4) For each document  $d$  and for each pair of words in the article  $b = (w_{i1}, w_{i2}), b \in B$ ,
  - (a) Choose a topic  $z_i \sim \theta_d z_i \sim \theta_d$
  - (b) Choose an emotional label  $l_i \sim \pi_{z_i}$
  - (c) Choose a pair of words  $b_i \sim \varphi_{z_i, l_i}$
  - (d) Choose a timestamp  $t_i \sim \psi_{z_i, l_i}$

As shown in Figure 2, word pairs in a document may belong to different timestamps in the text generation process of the TSTS topic model. In theory, all the content of an article such as words and topics should belong to the same timestamp. Also, the introduction of the time factor into the topic model will affect the topic homogeneity of an article. However, the default time factor of the TSTS model in the topic model will not affect the homogeneity of the text. So, it is assumed that the time factor in the paper has no weight. Based on the TOT and the group topic (GT) model, the superparameter  $\mu$  is introduced into TSTS to balance the interaction of time and words in document generation. The parameters' explanation of the TSTS model is shown in Table 1.

**3.3. Model Deduction.** According to the Bayesian network structure diagram of the TSTS model, the polynomial distribution  $\theta$  of the topic, the distribution  $\pi$  of sentiment with the topic, the correlation distribution  $\phi$  of word pairs with <topic, sentiment>, and the correlation distribution  $\psi$  of time with <topic, sentiment> can be calculated according to the superparameters  $\alpha, \beta, \gamma$ , and  $\mu$ . Then, Gibbs sampling is done that can ensure the convergence of the TSTS model under enough iteration times. And, each word in the document is assigned the topic and sentiment that are most suitable for the facts.

According to the principle of Bayesian independence, the joint probability of word pair, topic, sentimental polarity, and timestamp is given as follows:

$$p(\mathbf{b}, \mathbf{t}, \mathbf{l}, \mathbf{z} | \alpha, \beta, \gamma, \mu) = p(\mathbf{b} | \mathbf{l}, \mathbf{z}, \beta) \cdot p(\mathbf{t} | \mathbf{l}, \mathbf{z}, \mu) \cdot p(\mathbf{l} | \mathbf{z}, \gamma) \cdot p(\mathbf{z} | \alpha), \quad (1)$$

TABLE 1: Explanation of parameters.

|             |   |
|-------------|---|
| $D$         | Number of documents   |
| $V$         | Vocabulary size   |
| $T$         | Number of topics  |
| $S$         | Number of sentiment polarity  |
| $H$         | Number of timestamps  |
| $M$         | Number of word pairs  |
| $B$         | Set of word pairs   |
| $B$         | Word pairs, $b = (w_{i1}, w_{i2})$  |
| $W$         | Word  |
| $T$         | Time  |
| $Z$         | Topic   |
| $L$         | Sentiment polarity label  |
| $\Theta$    | $[\theta_d]$ : polynomial distribution of topics  |
| $\Phi$      | $[\varphi_{z,l}]$ : $T \times S \times V$ matrix, word pairs' distribution                                  |
| $\Pi$       | $[\pi_z]$ : $T \times S$ matrix, sentiment distribution   |
| $\Psi$      | $[\psi_{z,t}]$ : $T \times S \times H$ matrix, time distribution  |
| $\alpha$    | Dirichlet prior parameters of $\Theta$  |
| $\gamma$    | Dirichlet prior parameters of $\pi$   |
| $\beta$     | Asymmetric Dirichlet prior parameters of $\Phi$   |
| $\mu$       | Dirichlet prior parameters of $\psi$  |
| $n_d$       | The number of word pairs in document $d$  |
| $n_{d,j}$   | The number of word pairs for topic $j$ in document $d$  |
| $n_j$       | The number of word pairs assigned as topic $j$  |
| $n_{j,k}$   | The number of word pairs assigned as topic $j$ and sentiment polarity $k$                                   |
| $n_{i,j,k}$ | The number of word pair $b_i$ is assigned to the topic $j$ and sentiment polarity $k$                       |
| $n_{j,k,h}$ | The number of word pair $b_i$ is assigned to the topic $j$ and sentiment polarity $k$ when timestamp is $h$ |
| $n^{-p}$    | The number of word pairs in the current document except for the $p$ position                                |

where the parameters are independent such as word pairs  $\mathbf{b}$  and parameters  $\alpha, \gamma$ , and  $\mu$ , timestamps  $\mathbf{t}$  and parameters  $\alpha, \gamma$ , and  $\beta$ , sentiment polarity  $\mathbf{l}$  and parameters  $\alpha, \mu$ , and  $\beta$ , and topic words  $\mathbf{z}$  and parameters  $\beta, \gamma$ , and  $\mu$ . Therefore, the joint distribution in the equation can be obtained by calculating the four parts on the right side of the equation.

Given the sentiment polarity label of specific topic features, the distribution of  $\mathbf{b}$  can be regarded as a polynomial distribution. Based on the premise of topic words  $z_i$  and  $l_i$ ,  $b_i$  is generated by  $N$  times with the probability  $p(\mathbf{b} | \mathbf{l}, \mathbf{z})$  at each time. Given that word pairs are independent of each other, we can obtain

$$p(\mathbf{b} | \mathbf{l}, \mathbf{z}, \beta) = \prod_{i=1}^N p(b_i | z_i, l_i) = \prod_{i=1}^N \beta \cdot b_i. \quad (2)$$

Superparameters are the representation parameters of the framework in the machine learning model [28], such as the number of classes in the clustering method or the number of topics in the topic model. In the Bayesian network, the distribution and density function of  $\theta$  are denoted as  $H(\theta)$  and  $h(\theta)$ , respectively. They are regarded as the prior distribution function and the prior density function, respectively, which are collectively referred to as the prior distribution. If the distribution of  $\theta$  is obtained after sampling, it is called the posterior distribution. Based on the conjugate property of Dirichlet~multinomial, when the



parameters in the population distribution conform to the distribution law of polynomial (Multinomial), the conjugate prior distribution conforms to the following distribution:

$$\text{Dir}(\boldsymbol{\theta} | \boldsymbol{\alpha}) + \text{Mult}(\boldsymbol{\delta}) = \text{Dir}(\boldsymbol{\theta} | \boldsymbol{\alpha} + \boldsymbol{\delta}). \quad (3)$$

For the general text model, the discretized Dirichlet distribution and multinomial distribution are as follows:

$$\text{Dir}(\mathbf{b} | \boldsymbol{\beta}) = \frac{\Gamma(\sum_{j=1}^T \boldsymbol{\beta})}{\prod_{j=1}^T \Gamma(\boldsymbol{\beta})} \prod_{j=1}^T n_j, \quad (4)$$

$$\text{Mult}(n | \mathbf{b}, N) = \binom{N}{n} \prod_{j=1}^T n_j, \quad (5)$$

where  $i, j, k$ , and  $h$  represent the iteration times of word pairs, topic, sentiment, and timestamp in the modeling process, respectively. Since the distribution of  $p(\mathbf{b} | \mathbf{l}, \mathbf{z}, \boldsymbol{\beta})$  follows the Dirichlet distribution, this paper introduces  $\varphi$  for  $p(\mathbf{b} | \mathbf{l}, \mathbf{z}, \boldsymbol{\beta})$ . It can be obtained by integrating  $\varphi$ :

$$\begin{aligned} p(\mathbf{b} | \mathbf{l}, \mathbf{z}, \boldsymbol{\beta}) &= \int p(\mathbf{b} | \mathbf{l}, \mathbf{z}, \varphi) \cdot p(\varphi | \boldsymbol{\beta}) d\varphi \\ &= \left( \frac{\Gamma(V\boldsymbol{\beta})}{\Gamma(\boldsymbol{\beta})^V} \right)^{T \cdot S} \prod_j \prod_k \frac{\prod_i \Gamma(n_{i,j,k} + \boldsymbol{\beta})}{\Gamma(n_{j,k} + V\boldsymbol{\beta})}. \end{aligned} \quad (6)$$

To estimate the posterior parameters  $\varphi$  in the formula, we can combine with the Bayes formula and the conjugate property of Dirichlet~multinomial. The distribution of the posterior parameters can be obtained as follows:

$$p((\varphi | \mathbf{l}, \mathbf{z}, \boldsymbol{\beta})) \propto \text{Dir}(\varphi | n_{i,j,k} + \boldsymbol{\beta}). \quad (7)$$

Given that the expectation of the Dirichlet distribution is  $E(\text{Dir}(\boldsymbol{\varepsilon})) = \boldsymbol{\varepsilon}_i / \sum_j \boldsymbol{\varepsilon}_j$ , so the calculated parameters are estimated by the known posterior parameter distribution expectation. The estimated results are shown in equation (7). Similarly, for  $p(\mathbf{t} | \mathbf{l}, \mathbf{z}, \boldsymbol{\mu})$ ,  $\psi$  is introduced. By integrating  $\psi$ , it can be obtained as follows:

$$p(\mathbf{t} | \mathbf{l}, \mathbf{z}, \boldsymbol{\mu}) = \left( \frac{\Gamma(H\boldsymbol{\mu})}{\Gamma(\boldsymbol{\mu})^H} \right)^{T \cdot S} \prod_j \prod_k \frac{\prod_h \Gamma(n_{j,k,h} + \boldsymbol{\mu})}{\Gamma(n_{j,k} + H\boldsymbol{\mu})}. \quad (8)$$

For  $p(\mathbf{l} | \mathbf{z}, \boldsymbol{\gamma})$ ,  $\pi$  is introduced. By integrating  $\pi$ , it can be obtained as follows:

$$p(\mathbf{l} | \mathbf{z}, \boldsymbol{\gamma}) = \left( \frac{\Gamma(\sum_k \boldsymbol{\gamma}_k)}{\prod_k \Gamma(\boldsymbol{\gamma}_k)} \right)^T \prod_j \frac{\prod_k \Gamma(n_{j,k} + \boldsymbol{\gamma}_k)}{\Gamma(n_j + \sum_k \boldsymbol{\alpha}_k)}. \quad (9)$$

For  $p(\mathbf{z} | \boldsymbol{\alpha})$ ,  $\theta$  is introduced. By integrating  $\theta$ , it can be obtained as follows:

$$p(\mathbf{z} | \boldsymbol{\alpha}) = \left( \frac{\Gamma(\sum_j \boldsymbol{\alpha}_j)}{\prod_j \Gamma(\boldsymbol{\alpha}_j)} \right)^D \prod_d \frac{\prod_j \Gamma(n_{d,j} + \boldsymbol{\alpha}_j)}{\Gamma(n_d + \sum_j \boldsymbol{\alpha}_j)}. \quad (10)$$

The TSTS model can estimate the posterior distribution after estimated values  $z$  and  $s$  have been obtained by sampling calculations. Then, the calculated equations (2)–(6) are

brought into equation (1). Combining with the nature of Gamma function, the conditional distribution probability in Gibbs sampling can be obtained:

$$\begin{aligned} p(s_p = k, z_p = j | \mathbf{b}, \mathbf{t}, \Gamma^{-P}, \mathbf{z}^{-P}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\mu}) \\ \propto \frac{n_{d,j}^{-P} + \boldsymbol{\alpha}_j}{n_d^{-P} + \sum_j \boldsymbol{\alpha}_j} \cdot \frac{n_{w_p,j,k}^{-P} + \boldsymbol{\beta}}{n_{j,k}^{-P} + V\boldsymbol{\beta}} \cdot \frac{n_{j,k}^{-P} + \boldsymbol{\gamma}_k}{n_j^{-P} + \sum_k \boldsymbol{\gamma}_k} \cdot \frac{n_{j,k,t_p}^{-P} + \boldsymbol{\mu}}{n_{j,k}^{-P} + H\boldsymbol{\mu}}. \end{aligned} \quad (11)$$

In order to simplify equation (6), the superparameter  $\mu = 1/n_d$  is introduced. When the superparameters  $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}$ , and  $\boldsymbol{\gamma}$  are given, the set B of the word pair, the corresponding topic  $z$ , and sentiment label  $l$  can be used to infer the parameters  $\varphi, \theta, \pi$ , and  $\psi$  based on Bayes' rule and Dirichlet conjugate properties:

$$\begin{aligned} \varphi_{j,k,i} &= \frac{n_{i,j,k} + \boldsymbol{\beta}}{n_{j,k} + V\boldsymbol{\beta}}, \\ \theta_{d,j} &= \frac{n_{d,j} + \boldsymbol{\alpha}_j}{n_d + \sum_j \boldsymbol{\alpha}_j}, \\ \pi_{j,k} &= \frac{n_{j,k} + \boldsymbol{\gamma}_k}{n_j + \sum_k \boldsymbol{\gamma}_k}, \\ \psi_{j,k,h} &= \frac{n_{j,k,h} + \boldsymbol{\mu}}{n_{j,k} + H\boldsymbol{\mu}}. \end{aligned} \quad (12)$$

## 4. Experiment Analysis

**4.1. Data Collection.** In order to verify the TSTS model proposed in this paper, the four hot events are randomly selected from the trending searches of Sina Weibo in 2019. And, the comments of four events are regarded as the experimental datasets. The four datasets selected are "Military parade in National Day," "The assault on a doctor," "Hong Kong's event," and "Garbage sorting in Shanghai." The comments are extracted from the Sina social network platform. In the original datasets, there are some meaningless words in the microblog text, such as stop words, interjections in tone, punctuation marks, and numeric expressions. Before text modeling, the word segmentation package in Python is used to process the experimental initial dataset. In addition, considering that comments on social networks are relatively new, the fashionable expressions in the social network are collected and added to the customized dictionary. So, these emerging words can be identified as far as possible and replaced with normal expressions. In addition, there are some useless words in the text, such as URL links and numbers, which can be filtered by regular expressions. Finally, a total of 14288 experimental data in four events are obtained. The description of four datasets is shown in Table 2.

**4.2. Sentiment Dictionary.** The words or phrases in the sentiment dictionary have obvious sentiment tendency,

TABLE 2: Experiment datasets.

| The dataset (the number of comments) | Number of words in per microblog |              | Vocabulary size |              |
|--------------------------------------|----------------------------------|--------------|-----------------|--------------|
|                                      | Initial                          | Pretreatment | Initial         | Pretreatment |
| The dataset 1 (3562)                 | 134                              | 102          | 9789            | 6319         |
| The dataset 2 (3527)                 | 127                              | 94           | 9736            | 6242         |
| The dataset 3 (3617)                 | 131                              | 100          | 9780            | 6301         |
| The dataset 4 (3582)                 | 128                              | 96           | 9742            | 6254         |
| Average                              | 130                              | 98           | 9762            | 6279         |

which can be divided into positive and negative words. The sentiment dictionary in this paper has two major roles. On the one hand, we can identify sentiment polarity words and distinguish topic features and sentiment words. On the other hand, combining with sentiment prior information to make the model more accurate in judging the sentiment polarity of the text. Given that sentiment polarity words can reflect users' sentiment tendency, it is of great significance to analyze the sentiment orientation of the text.

At present, there are two major Chinese sentiment dictionaries: NTU and HowNet. The former dictionary contains 2812 positive words and 8276 negative words. The latter contains about 5,000 positive words and 5,000 negative words. Based on HowNet and the classification of sentiment polarity [29, 30], this paper constitutes the sentiment dictionary of the TSTS model evaluation experiment, as shown in Table 3.

**4.3. Parameter Setting.** In this paper, the Gibbs algorithm is used to sample the TSTS model and estimate four posterior parameters. According to the parameter setting in the traditional topic model, the superparameters are set as follows. First, the superparameter  $\alpha$  is set as  $50/K$ , and  $K$  is the number of topics extracted. Second,  $\beta$  is set as 0.01. Third,  $\gamma$  is set as  $(0.05 \times \text{AVE})/S$ . AVE stands for the average length of articles, that is, the average number of words in the microblog in this experiment, and  $S$  stands for the total number of polar tags. Finally,  $\mu$  is set as  $1/n_d$ .

**4.4. Evaluation Indicator.** For the extraction of topic features, perplexity is used as an evaluation indicator to measure the predictive power of unknown data in the process of model modeling. Also, the lower perplexity means better efficiency. The calculation formula of the perplexity is as follows:

$$\text{perplexity} = P(\bar{D}_t | \mathcal{M}) = \exp \left\{ -\frac{\sum_{d=1}^{D^t} \log P(\bar{b}_d^t | \mathcal{M})}{\sum_{d=1}^{D^t} \bar{N}_d^t} \right\}, \quad (13)$$

where  $\bar{D}_t = \left\{ \bar{b}_d^t \right\}_{d=1}^{D^t}$  represents an unknown dataset with the timestamp  $t$ .

$$P(\bar{b}_d^t | \mathcal{M}) = \prod_{n=1}^{\bar{N}_d^t} \prod_{l=1}^L \prod_{i=1}^T P(\bar{b}_{d,n} | l, z) P(z | l) P(l), \quad (14)$$

TABLE 3: Classification of sentiment words.

| Sentiment labels | Happy | Surprise | Sad  | Angry |
|------------------|-------|----------|------|-------|
| Vocabulary size  | 2467  | 276      | 3025 | 1897  |

where  $\bar{b}_d^t$  represents the vector set of word pairs in text  $d$ ,  $\bar{N}_d^t$  represents the number of word pairs in  $\bar{b}_d^t$ , and  $P(\bar{b}_d^t | \mathcal{M})$  represents the direct possibility of training corpus, and the formula is as follows:

$$P(\bar{b}_d^t | \mathcal{M}) = \prod_{i=1}^V \left( \sum_{l=1}^L \sum_{z=1}^T \varphi_{l,z,i} \cdot \theta_{d,l,z} \cdot \pi_{d,l} \right)^{\bar{N}_{di}^t}. \quad (15)$$

For sentiment segmentation, the sentiment judgment from the perspective of the document is used as the evaluation index, which is based on the sentiment polarity label in the sentiment dictionary. For the documents in this experiment, the positive and negative sentiment of a document can be judged. This paper adopts the consistency test method to mark the sentiment labels [31].

## 5. Results

**5.1. Extraction of Topics.** The primary task of the TSTS model is to extract topic features. As an extension of the topic-sentiment mixed model, the assessment is to judge whether the extracted topic features are reasonable and accurate. Before extracting topic features from text modeling, it is necessary to determine the number of topics to be extracted and the iteration times of Gibbs sampling. For the effective evaluation of topic discovery, the degree of perplexity is used as the measurement index in the paper. The lower the perplexity is, the better the fitting effect of the model is. Taking dataset 1 as an example, the simulation results are shown in Figure 3.

Based on the experimental results shown in Figure 3, the number of topics was set 20 in the subsequent experiments. In addition, we can calculate the perplexity of three models with the change of the iterations. By comparing the experimental results of TSTS and LDA, it can be found that the effect of TSTS is always better than LDA, and the degree of perplexity decreases with the increase of iteration. That indicates that the topic discovery ability of TSTS gradually improves, mainly because the TSTS model incorporates the word pairs to alleviate the sparse matrix of LDA for short texts. By comparing the experimental results of TSTS and BTM, it can be found that TSTS was better than BTM when the number of iterations increases. However, as the number

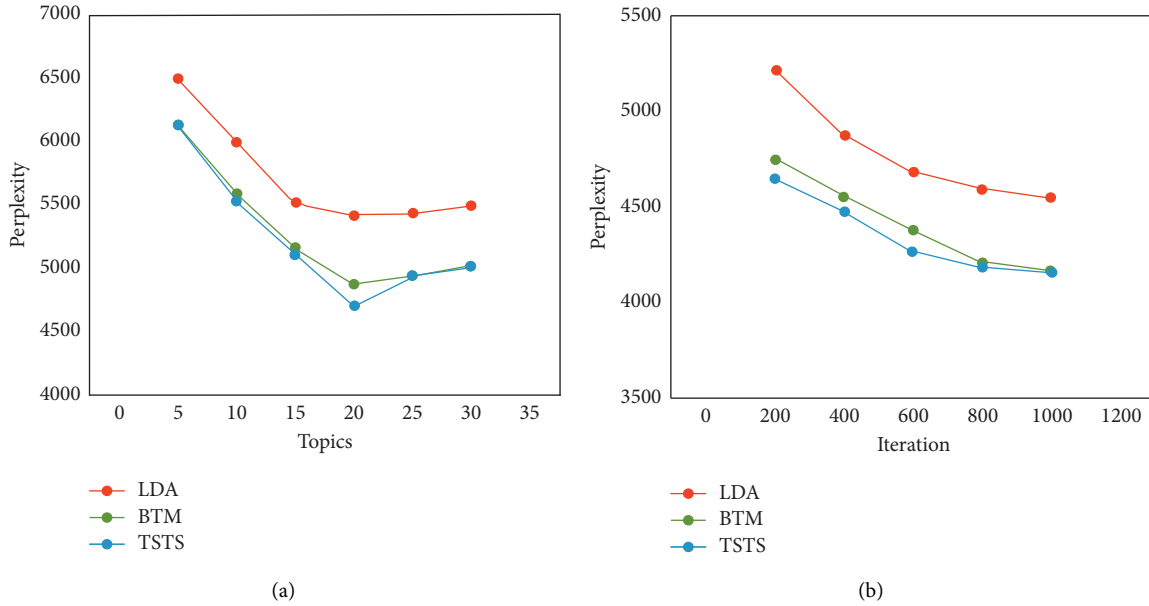


FIGURE 3: The relationship of the perplexity with the number of (a) topics and (b) iterations.

of iterations increased, the gap of both models became smaller. This is because the word pair of BTM is used for the whole corpus. When the number of iterations is small, the proportion of noise words is relatively large, resulting in poor quality of topic words. In addition, the sentiment layer is integrated into TSTS, and the error generated in the sentiment estimation will affect the next iteration. Although TSTS is worse than BTM when there are more iterations, the effect of TSTS can still be balanced with BTM. Therefore, during the extraction of topic features, the number of topics and iterations can be set as 20 and 600.

**5.2. Sentiment Polarity.** The information related to sentiment polarity is provided in accordance with the topic and sentiment polarity of words. The sentiment distribution of topics extracted from the TSTS model is shown in Figure 4. In addition, JST and ASUM are introduced as the comparison to measure the effect of sentiment recognition of the TSTS model. Each document has a binary sentiment label, such as positive or negative sentiment. Taking dataset 2 “The attack on the doctor” as an example, the result is shown in Figure 4. The number of topics is set to 5 at the beginning of the experiment. With the refinement of granularity, the performance of the TSTS model increases. Compared with JST and ASUM, the curve of the TSTS model changes greatly considering the topic and sentiment relationship among word pairs of the document. The change curve of the JST model shows a steady upward trend, and the identification efficiency of ASUM is low. That is because ASUM has strict assumptions, and the increase in the number of topics will cause the decentralization of topics and sentiments, which has a great negative impact on the overall performance of the model. The overall effect of the TSTS model was slightly better than JST and ASUM, but the effect decreased slightly after the number of topics increased to 20. This is because the

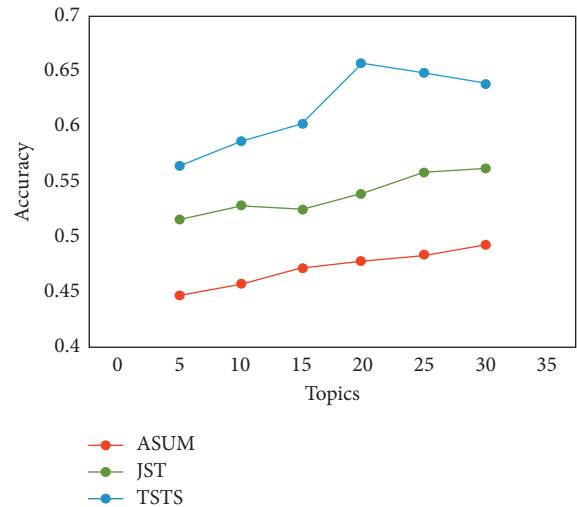


FIGURE 4: Accuracy of sentiment polarity judgment.

data collected in the dataset are limited, and the number of topics has been set to discretize the word distribution. Thus, the judgment of sentiment polarity is affected. The sentiment label classification of documents is compared under different topics, and the result of the TSTS model is better than JST and ASUM.

With the increase of topics, the recognition performance of the topic model has some fluctuations. But, the TSTS model was always better than JST and ASUM. When the number of obtained topics and iterations is set as 20 and 600, the TSTS is the best model in topic detection. When the number of topics in the four datasets was set as 20, the accuracy of sentiment polarity judgment is shown in Table 4.

From Table 4, the TSTS model is better than JST and ASUM in judging the sentiment polarity of documents. This

TABLE 4: Accuracy of sentiment polarity judgment.

|               | ASUM   | JST    | TSTS   |
|---------------|--------|--------|--------|
| The dataset 1 | 0.4763 | 0.5427 | 0.6348 |
| The dataset 2 | 0.4617 | 0.5398 | 0.6599 |
| The dataset 3 | 0.4832 | 0.5461 | 0.6475 |
| The dataset 4 | 0.4841 | 0.5294 | 0.6522 |

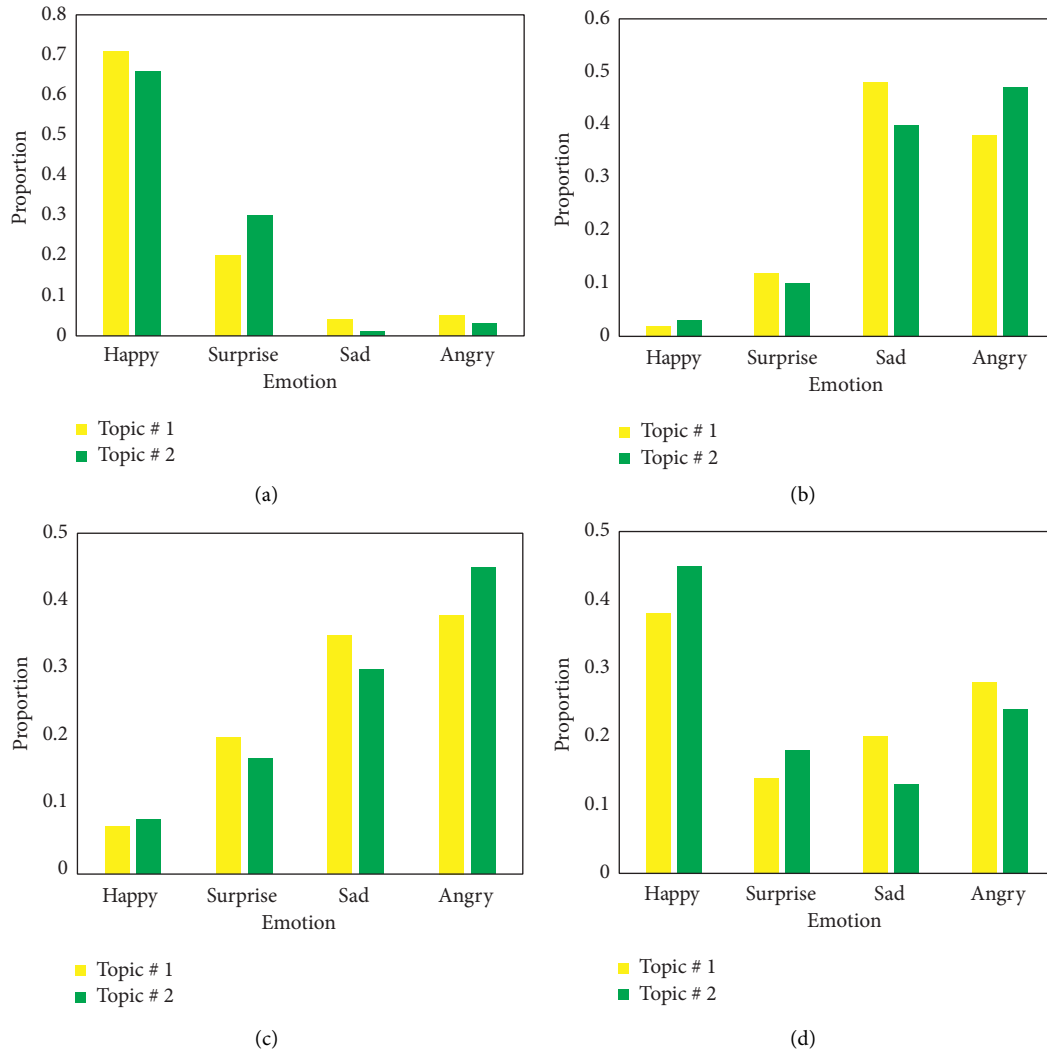


FIGURE 5: Sentiment distribution in four datasets. (a) Military parade in National Day. (b) The assault on a doctor. (c) Hong Kong’s event. (d) Garbage sorting in Shanghai.

is because sentiment polarity depends on the performance of topic discovery in the previous stage. In this experiment, the effect of JST and ASUM was exactly opposite. The difference was caused by the length of the original document, which also indirectly verified the effectiveness of the TSTS model.

From Figure 5, it can be seen that the proportion of positive sentiment is significantly higher than the other sentiments in the dataset “Military parade in National Day” and the dataset “Garbage sorting in Shanghai,” which is consistent with the sentiment tendency of users in the social network. For the second dataset “The assault on a doctor,” the two kinds of negative sentiment polarities of topic #1 and

topic #2 were compared. Topic #1 is more likely to be sad sentiment, while topic #2 is more likely to be angry sentiment. Topic #1 reflects the statement of the event, and topic #2 represents the follow-up discussion of the event.

5.3. *Topic and Sentiment Evolution.* The curves of topic features extracted from four datasets through the TSTS model are shown in Figure 6. Taking dataset 2 as an example, the topic curve conforms to the evolution law of social and abrupt events. The two curves represent the trend of feature words over time in topic #1 and topic #2. Topic #1 is the

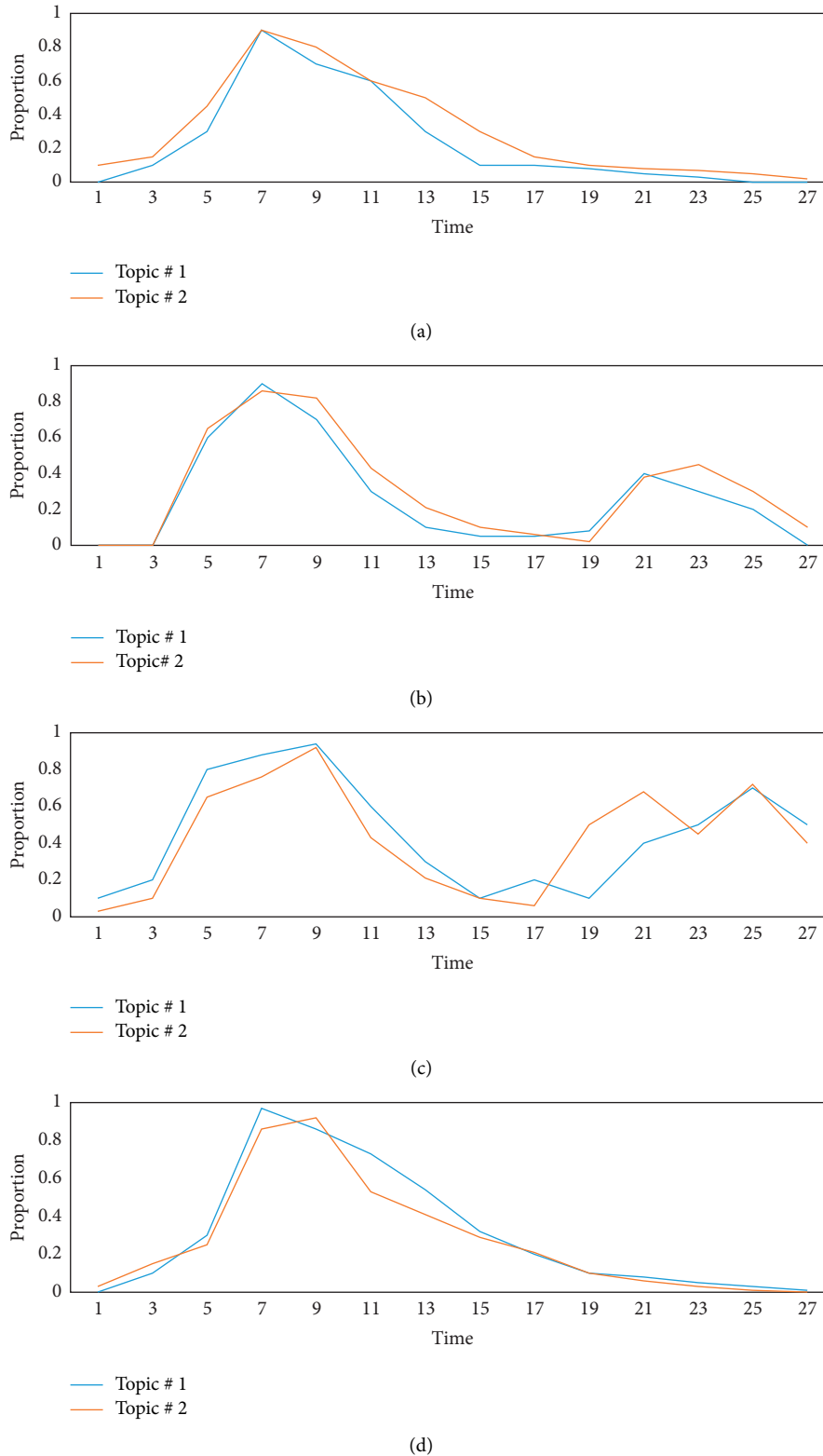
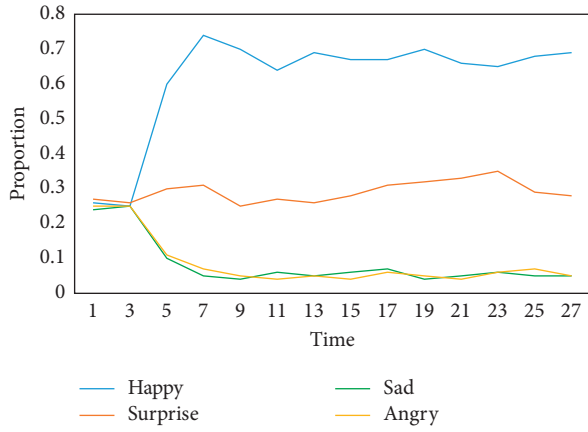


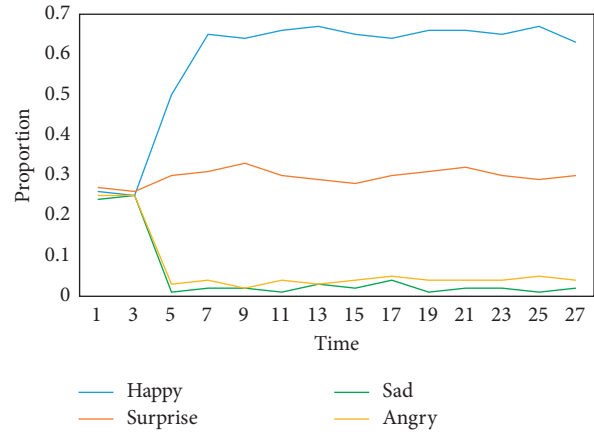
FIGURE 6: The changing of topics in four datasets. (a) Military parade in National Day. (b) The assault on a doctor. (c) Hong Kong's event. (d) Garbage sorting in Shanghai.

statement about the case itself. From the beginning of the event, the amount of discussion about the event on the social network rose sharply and then gradually declined. Topic #2

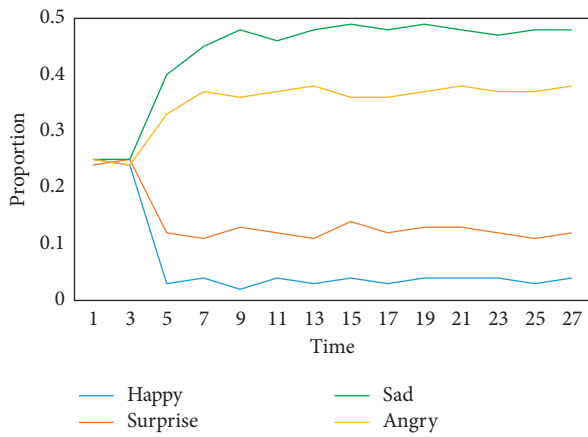
is a discussion on the development of the case, which has caused the second hot discussion again. The time is not consistent when two curves reach the high peak. The peak



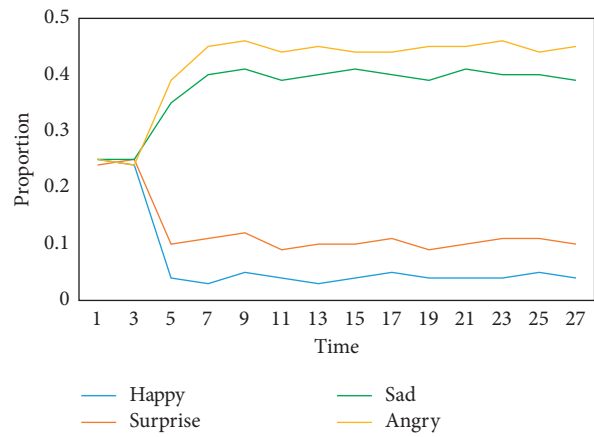
(a)



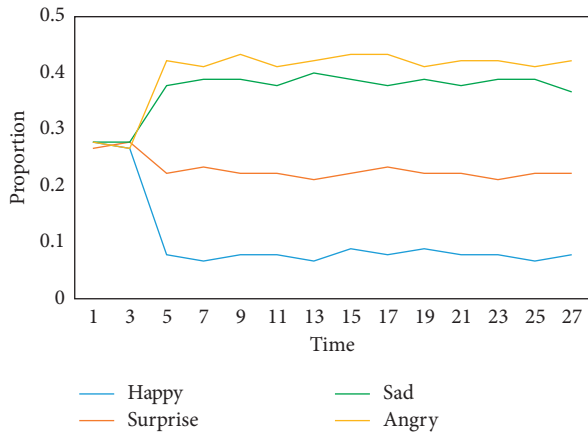
(b)



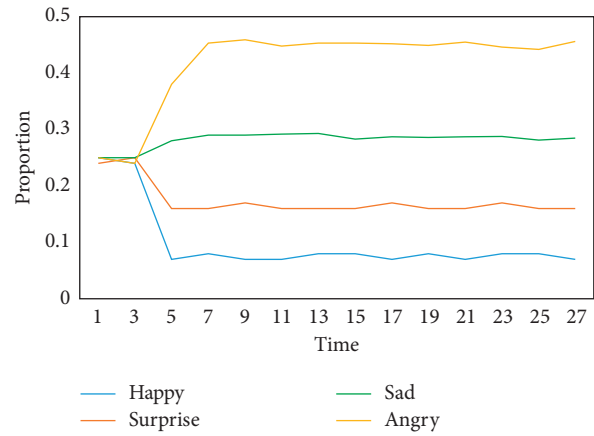
(c)



(d)



(e)



(f)

FIGURE 7: Continued.

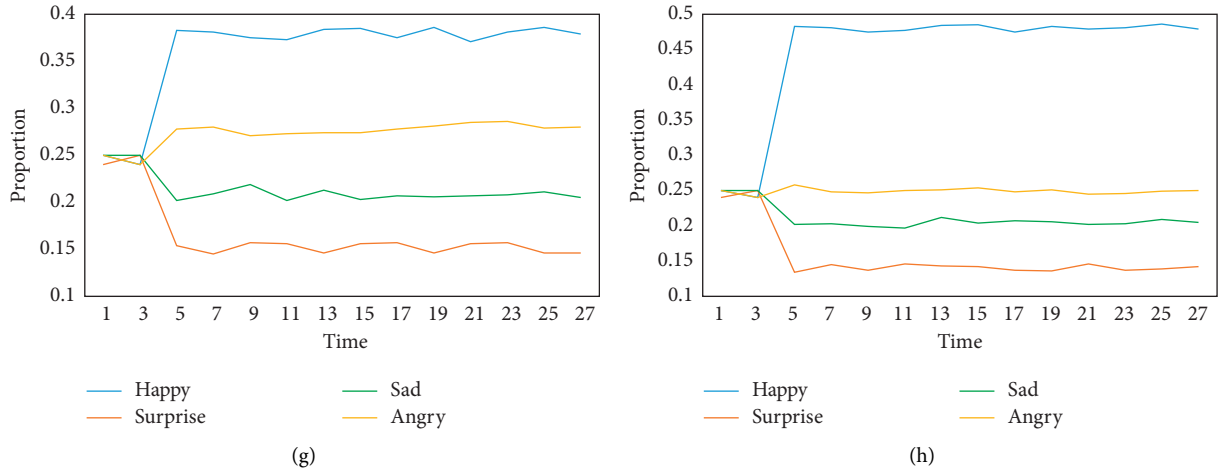


FIGURE 7: The changing of sentiment in four datasets. (a) Military parade in National Day topic #1. (b) Military parade in National Day topic #2. (c) The assault on a doctor topic #1. (d) The assault on a doctor topic #2. (e) Hong Kong’s event topic #1. (f) Hong Kong’s event topic #2. (g) Garbage sorting in Shanghai topic #1. (h) Garbage sorting in Shanghai topic #2.

value of topic #2 curve is lower than that of topic #1, which reflects the discussion of the same event will fade over time. Even if there is a new topic, discussion of a new topic is far lower than the beginning of the event. Meanwhile, similar results can be verified in other three datasets.

The proportion of the sentiment polarity in the four datasets is shown in Figure 7. Since the sentiment polarity proportion is measured, the four sentiment polarities are balanced distributed before the occurrence of the events. After the event occurred, the polarity of positive and negative sentiment began to change toward the two extremes. Among the four datasets, the positive sentiment was higher than the negative sentiment in the first dataset “National Day military parade event” and the fourth dataset “Shanghai garbage classification event,” which also conforms to the social sentiment of events. In addition, it can be found that the difference among the four sentiment labels is large in the initial stage, and the distribution of sentiment labels becomes stable in the later stage. It has proved that the second report of social events does not cause the heat for the first time. But, the sentiment tendency judgment in social networks will not decline sharply with the reduction of discussion, which can be proved in topic #2 of the second dataset “The assault on a doctor.” Given that the topic features come from the background of a corpus and contain a lot of noise words, the relative position of four curves is closer in terms of sentiment polarity evolution. However, there is still a gap between government feelings, which is different from the average distribution of sentiment polarity at the beginning of the event.

## 6. Discussion

From the perspective of theoretical significance, this paper extends the LDA model to some extent. First, in view of the sparse matrix caused by the short text, word pairs are introduced to replace a single word for text generation according to BTM. Based on the hypotheses of JST and

ASUM, the sentiment layer is introduced to form the Bayesian network structure, and the word pair is limited to the same sentiment polarity distribution. Second, in order to realize dynamic analysis and text homogeneity, the timestamp and the corresponding superparameter are introduced to alleviate the problem of the word order in the text generation. Third, this research combines behavioral experiments, big data mining, mathematical modeling, and imitating to promote the research expansion of new situations and new methods.

From the perspective of practical significance, this paper is of great value in tracking and monitoring topics of public opinion in social networks. The online public opinions of hot events can be monitored, which contributes to accurately judge social events and make emergency decisions for government or departments. In addition, this paper analyzed the evolution, response, and governance of public opinion, which is conducive to understand the formation mechanism and the collaborative evolution of public opinion. Meanwhile, the use of public opinion information can detect and screen information, prevent the spread of rumor, and scientifically formulate the mechanism of utilization to effectively reduce the loss risk.

## 7. Conclusions

In the context of the mobile social network, the number of short texts is growing explosively. In order to extract information from massive short texts quickly and monitor public opinions, the TSTS model is proposed in the study based on LDA, BTM, JST, ASUM, and TOT. From the experimental results, the TSTS model achieves good performance. In terms of topic feature extraction, the degree of perplexity of TSTS is always lower than LDA. Moreover, although the degree of perplexity is slightly higher than BTM with the increase of iteration times, it can maintain the balance with BTM. In the sentiment analysis, the effect of TSTS was significantly better than JST and ASUM. Finally,

the TSTS model incorporating the time factor can determine the change trend of the topic and sentiment.

There are still some shortcomings in this paper. Firstly, for the extraction of topic feature words, the global topic level can be added to the topic layer of the TSTS model to filter the common topic words. Secondly, in sentiment polarity judgment, the sentiment labels are manually marked based on prior knowledge. However, the sentiments are extremely rich and changeable. In the future research, the Bayesian network and entity theory can be used to judge sentiment bias.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported in part by the Project of National Science and Technology Department (2018YFF0213102), Public Projects of Zhejiang Province (LGF18G010003, LGF19G010002, and LGF20G010002), Science and Technology Project of Zhejiang Province (2020C01158), and First Class Discipline of Zhejiang - A (Zhejiang Gongshang University - Statistics).

## References

- [1] CNNIC, "Statistical report on internet development in China," CNNIC, Beijing, China, 2019, [http://www.cac.gov.cn/2019-08/30/c\\_1124938750.htm](http://www.cac.gov.cn/2019-08/30/c_1124938750.htm).
- [2] M. Ingawale, A. Dutta, R. Roy, and P. Seetharaman, "Network analysis of user generated content quality in wikipedia," *Online Information Review*, vol. 37, no. 4, pp. 602–619, 2013.
- [3] Q. Gao, Y. Tian, and M. Tu, "Exploring factors influencing Chinese user's perceived credibility of health and safety information on Weibo," *Computers in Human Behavior*, vol. 45, pp. 21–31, 2015.
- [4] M. Steyvers and T. Griffiths, "Probabilistic Topic Models," *Handbook of Latent Semantic Analysis*, Psychology Press, vol. 427, no. 7, pp. 424–440, New York, NY, USA, 2007.
- [5] J.-F. Yeh, Y.-S. Tan, and C.-H. Lee, "Topic detection and tracking for conversational content by using conceptual dynamic latent Dirichlet allocation," *Neurocomputing*, vol. 216, pp. 310–318, 2016.
- [6] X. Zhou and L. Chen, "Event detection over twitter social media streams," *The VLDB Journal*, vol. 23, no. 3, pp. 381–400, 2014.
- [7] Y. J. Du, Y. T. Yi, and X. Y. Li, "Extracting and tracking hot topics of micro-blogs based on improved Latent Dirichlet allocation," *Engineering Applications of Artificial Intelligence*, vol. 87, pp. 1–13, 2020.
- [8] H. J. Kim, Y. K. Jeong, Y. Kim et al., "Topic-based content and sentiment analysis of Ebola virus on twitter and in the news," *Journal of Information Science*, vol. 42, no. 6, pp. 763–781, 2016.
- [9] B. Subeno, R. Kusumaningrum, and F. Farikhin, "Optimisation towards latent Dirichlet allocation: its topic number and collapsed gibbs sampling inference process," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 5, pp. 3204–3213, 2018.
- [10] H. Park, T. Park, and Y.-S. Lee, "Partially collapsed gibbs sampling for latent Dirichlet allocation," *Expert Systems with Applications*, vol. 131, pp. 208–218, 2019.
- [11] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.
- [12] X. Cheng, X. Yan, Y. Lan, and J. Guo, "BTM: topic modeling over short texts," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 12, pp. 2928–2941, 2014.
- [13] J. Rashid, S. M. A. Shah, and A. Irtaza, "Fuzzy topic modeling approach for text mining over short text," *Information Processing & Management*, vol. 56, no. 6, pp. 1–19, 2019.
- [14] H.-Y. Lu, N. Kang, Y. Li, Q.-Y. Zhan, J.-Y. Xie, and C.-J. Wang, "Utilizing recurrent neural network for topic discovery in short text scenarios," *Intelligent Data Analysis*, vol. 23, no. 2, pp. 259–277, 2019.
- [15] L. Zhu, H. Xu, Y. Xu et al., "A joint model of extended LDA and IBTM over streaming Chinese short texts," *Intelligent Data Analysis*, vol. 23, no. 3, pp. 681–699, 2019.
- [16] M. Tang, J. Jin, Y. Liu et al., "Integrating topic, sentiment and syntax for modeling online product reviews: a topic model approach," *Journal of Computing and Information Science in Engineering*, vol. 19, no. 1, pp. 1–12, 2019.
- [17] R. K. Amplayo, S. Lee, and M. Song, "Incorporating product description to sentiment topic models for improved aspect-based sentiment analysis," *Information Sciences*, vol. 454–455, pp. 200–215, 2018.
- [18] L. Yao, Y. Zhang, B. Wei et al., "Concept over time: the combination of probabilistic topic model with Wikipedia knowledge," *Expert Systems with Applications*, vol. 60, pp. 27–38, 2016.
- [19] S. Park, W. Lee, and I.-C. Moon, "Associative topic models with numerical time series," *Information Processing & Management*, vol. 51, no. 5, pp. 737–755, 2015.
- [20] P. Lorenz-Spreen, F. Wolf, J. Braun et al., "Tracking online topics over time: understanding dynamic hashtag communities," *Computational Social Networks*, vol. 5, no. 1, pp. 1–18, 2018.
- [21] Y. He, C. Lin, W. Gao et al., "Dynamic joint sentiment-topic model," *ACM Transactions on Intelligent Systems & Technology*, vol. 5, no. 1, pp. 1–21, 2013.
- [22] L. Kuo and T. Y. Yang, "An improved collapsed Gibbs sampler for Dirichlet process mixing models," *Computational Statistics & Data Analysis*, vol. 50, no. 3, pp. 659–674, 2006.
- [23] Y. Papanikolaou, J. R. Foulds, T. N. Rubin et al., "Dense distributions from sparse samples: improved Gibbs sampling parameter estimators for LDA," *Statistics*, vol. 18, no. 62, pp. 1–58, 2015.
- [24] X. Zhou, J. Ouyang, and X. Li, "Two time-efficient Gibbs sampling inference algorithms for biterm topic model," *Applied Intelligence*, vol. 48, no. 3, pp. 730–754, 2018.
- [25] P. Bhuyan, "Estimation of random-effects model for longitudinal data with non ignorable missingness using Gibbs sampling," *Computational Statistics*, vol. 34, no. 4, pp. 1963–1710, 2019.
- [26] C. Lin, Y. He, R. Everson, and S. Ruger, "Weakly supervised joint sentiment-topic detection from text," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 6, pp. 1134–1145, 2012.



- [27] A. Daud and F. Muhammad, "Group topic modeling for academic knowledge discovery," *Applied Intelligence*, vol. 36, no. 4, pp. 870–886, 2012.
- [28] Z. Huang, J. Tang, G. Shan, J. Ni, Y. Chen, and C. Wang, "An efficient passenger-hunting recommendation framework with multi-task deep learning," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 7713–7721, 2019.
- [29] S. M. Mohammad, S. Kiritchenko, and X. Zhu, "NRC-Canada: building the state-of-the-art in sentiment analysis of tweets," in *Proceedings of the Seventh International Workshop on Semantic Evaluation Exercises (SemEval-2013)*, Springer, Atlanta, GA, USA, pp. 1–5, June 2013.
- [30] T. Chen, Q. Li, J. Yang, G. Cong, and G. Li, "Modeling of the public opinion polarization process with the considerations of individual heterogeneity and dynamic conformity," *Mathematics*, vol. 7, no. 10, p. 917, 2019.
- [31] S. A. Curiskis, B. Drake, T. R. Osborn, and P. J. Kennedy, "An evaluation of document clustering and topic modelling in two online social networks: twitter and Reddit," *Information Processing and Management*, vol. 57, no. 2, pp. 1–21, 2019.

## Research Article

# A Complexity Analysis of User Interaction with Hotel Robots

Lina Zhong <sup>1</sup>, Liyu Yang,<sup>1</sup> Jia Rong,<sup>2</sup> and Xiaonan Li<sup>1</sup>

<sup>1</sup>*Institute for Big Data Research in Tourism, School of Tourism Sciences, Beijing International Studies University, Chaoyang District, Beijing, China*

<sup>2</sup>*Department of Data Science & AI, Monash University, Melbourne, Australia*

Correspondence should be addressed to Lina Zhong; [zhonglina@bisu.edu.cn](mailto:zhonglina@bisu.edu.cn)

Received 28 March 2020; Accepted 12 May 2020; Published 30 May 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Lina Zhong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Service robots have been introduced to hotel industry in the past decade and received various feedback on their performance. To provide better service, one needs to understand how the hotel customers look at the service robots. Understanding their interests, motivation, and behaviors in human-robot interaction is the key to develop high-quality services and improve robot's performance. This is the first work to study human-robot interaction in hotels in China. Frequent pattern mining and social network analysis techniques are used in this work to find out useful suggestions to both hotel management and robot manufactory. Turning on and off lights, TV, curtain, and window screens are popular services that most of the hotel customers preferred during their stays. Service robots are also found to entertain customers to carry out repeated commands for fun or to kill time. Customers also showed various motivations to stay in hotel rooms by calling different commands.

## 1. Introduction

Rapid development of IT and AI makes it possible to have intelligent robots in people's daily life. In terms of functionality, robots could be classified as industry robots and service robots [1]. By a close look at the service robots, a further categorization can be applied for professional use or for personal/private use [2]. Service robots for professional use are designed to accomplish a particular task or serve a target group of users. For example, professional service robots include cleaning robots, sewer robots, inspection robots, demolition robots, underwater robots, robots in medical institutions, robots with disabilities such as auxiliary robots and wheelchair robots, delivery robots, guidance robots, gas station robots, firefighting robots and air defense robots, construction robots, agricultural robots, etc. Service robots for personal and private use include domestic (home) robots, entertainment robots, educational robots, and so on. These service robots have been developed to assist human beings to finish daily tasks [3]. Both service robots are designed to assist and interact with human beings to complete tasks.

Current research focuses on the design, application, and evaluation of service robots. In terms of design, how do designers design robots, robot interactions, and experiences based on customer needs and preferences? They can weaken or strengthen customers' perception of value, leading to positive or negative word of mouth [4]. Ivanov and Webster [5] focus on the design of friendly hospitality robots. Korea Institute of Science and Technology (KIST) designed elderly robot [6]. Lakshmi et al. [7] designed a hospital nursing drug delivery robot. In terms of applications, unlike industrial robots, the accuracy or speed of service robots is not always the most important aspect. Their function or purpose is not only for entertainment but also to provide help, guidance, treatment, education, and communication [3]. In terms of evaluation, Ivanov et al. [8] used a questionnaire to explore how Iranians view the hotel robots, especially what Iranian consumers think robots can do for them and what they want robots to do. Ivanov et al. [8] investigated the attitude of young Russians towards introducing robots in hotels. Tussiyadiah and Park [9] focused on customer evaluation of hotel service robots. There is still a lack of understanding of the content and effects of robot services, which is why the paper is written. The existing studies have not explored

much in the service content of the hotel robots or effectiveness of the human-robot interaction.

Human-robot Interaction was distinguished by Thrun [10] into two categories: direct and indirect interactions. Direct interaction assumes a type of two-way communication, which shows equal status between humans and robots. Indirect interaction assumes a type of one-way communication, where the robot acts according to the users' command and reacts to its users. This work focuses on indirect interaction with remote-control robots. The remote-control robot is controlled by a human from a distance (near or far), and the human can control the robot through a joystick or instruction [11].

This work aims to study the actual services requested by human to be implemented by hotel robots by analyzing historical data collected from the real hotel rooms equipped with service robots from 2017 to 2018. This work tried to answer questions such as (1) what are the most popular services the users requested from the hotel robots, (2) what are the motivations for humans to interact with hotel robots, and (3) how often do users give commands to hotel robots. Furthermore, the responsiveness of the hotel robots to the commands received is another interesting topic in this work.

The rest of the paper is organized as follows: Section 2 reviews the current research work on service robots in hotel industry and customer feedback on their performance; Section 3 introduces the data resource and collection process followed by the details of selected data analysis methods; the interesting user command patterns as the experimental results are presented and discussed in Section 4; Section 5 concludes the entire work together with limitation and possible future work.

## 2. Literature Review

To conduct the study of human-robot interactions, the related literature has been carefully reviewed to have a comprehensive understanding of the current work of service robot applications in relevant industry, especially in hotel industry, and the technology level and the performance of the current hotel service robots in real world.

*2.1. Robot Applications in Service Industry.* Service robots are designed to support and service humans through physical and social interactions [12]. For frontline service robots, Martins [13] defined them as follows: a service robot is a system-based autonomous and adaptive interface to communicate and provide services to customers. Service robots play different roles, such as pets, companions, secretaries (or subordinates), and bystanders, just like the relationship with humans [14, 15]. Service robots can significantly reduce labor costs, and they are more like "tools" of technology and employees than their replacements [16]. Martins [13] considered the following three attributes related to service design: representation, personification, and task orientation. The service robot can have a physical representation (such as Pepper) or only a virtual representation (such as Alexa). Therefore, we believe that virtual AI software that can run

autonomously and learn over time can also be classified as a service robot. The service robot can be designed as a humanoid robot (such as anthropomorphic robot) that simulates the appearance of a human (such as Sophia) or it can be designed as a nonhumanoid robot (such as a cleaning robot). Finally, service robots can perform cognitive analysis tasks according to basic computer functions (such as image analysis software assistants for medical diagnosis) or emotional social tasks (such as receiving robots).

Robots are widely used in the service industry. In terms of home education, South Korea has developed the world's first available e-learning home robot and demonstrated the future of robots as a new educational medium. The survey found that home robots are superior in promotion and application and are more able to focus students' attention and interests and improve their academic performance [17]. Ju et al. [18] designed an educational robot platform for the needs of robot education for elementary and middle school students. In terms of senior care, van Osch et al. [11] developed a home care robot called Rose, which is controlled from a long distance (8 km) to perform small tasks for the elderly. The Korea Institute of Science and Technology (KIST) has launched a new project to develop an elderly service robot named T-Rot at the Intelligent Robot Center [6]. Lee and Naguib [19] introduced the design and implementation of the next-generation elderly care robot "Home Mate" based on innovative commitment to sociality and reliability and extensive user research, which can be used for infotainment, video chat, games, and drug reminders. In terms of guidance, Bohus and Horvitz [20] introduced directional robots that can use natural language to interact with one or more participants and provide directions to offices, conference rooms, and other public areas within the building, such as kitchens, cafeterias, and bathrooms. Kanda et al. [21] developed a robot for a shopping mall designed to interact naturally with customers and provide shopping information emotionally. Linder et al. [22] experimentally tested the performance of tracking robots in dense places such as airport terminals for shopping malls. In terms of medicine, Mirheydar and Parsons [23] reviewed the application of robotic surgery in urology and pointed out that the safety of patients with robotic prostatectomy increases. Lakshmi et al. [7] introduced the design and manufacture of a drug delivery robot for hospitals to provide alternate care services. Hu et al. [24] designed a planner based on Gaussian process classification, which aims to promote the further integration of multifinger manipulators and medical image detection.

*2.2. Service Robots in Hotel Industry.* Service automation, artificial intelligence, and robotics offer huge opportunities for the hospitality industry [25]. Hotel robot functions include interacting with travelers and entertaining them, as well as physical assistance, such as transport, checking baggage, and guidance. [26]. Starwood has introduced a robot butler in their Aloft hotel mainly to provide conveniences to guest rooms [27]. Henn-na Hotel is the first hotel to hire hotel robots during the entire operation from check-

in to automatic baggage delivery, using robots to check in occupants and escort guests to their rooms. The robotic receptionist speaks Japanese or English, depending on guest preferences. It can set reservations for people, take them to their rooms, and adjust the temperature of their accommodation. In the room, guests can use voice commands to change the lights and ask questions about time or weather [28]. In collaboration with IBM, Hilton Worldwide is the first to launch the world's first robotic concierge service (using SoftBank's NAO robots) to gain knowledge from artificial intelligence systems and inform guests of local attractions, restaurants, hotel facilities, etc. [29]. The Las Vegas Wayne Hotel has announced that it will introduce Amazon Echo voice roll-up speakers in all rooms with Alexa digital assistants [30]. By the end of 2018, Alibaba's first future unmanned hotel "Flyzoo Hotel" has been put into operation [31].

From the demand side, economically, hotel robots can help hotels cope with seasonal employment and labor costs. Dirican [32] believes that artificial intelligence, especially the introduction of robotic services to hotels, greatly improves the efficiency of hotel services and reduces hotel operating costs, although some hotels with robots still consider investment costs to be high. With increasing labor costs, decreasing labor availability, decreasing robot costs, and increasing robot capabilities, many low-wage service jobs in the hotel industry are expected to be replaced [33]. The time spent on robot labor is less expensive than paying humans [34]. On the noneconomic side, using robots to complete some daily tasks may be considered by hotel managers as a good choice because it will save them most of the concerns of legal authorities, unions, and immigration offices such as seasonal layoffs. [12, 35]. Stringam and Gerdes [36] used a combination of quantitative and qualitative methods, pointing out that the self-service of hotels has not yet led to a significant reduction in labor. Due to hotel technology's labor savings, most hotels choose to redeploy employees to other tasks. It is improving customer service, not reducing staffing. Managing robots will be different than managing tangible goods. How robots, staff, and customers interact is an important focus for future travel and hotel managers [37].

From the supply side, Murphy et al. [38] discussed the design of robots for hotel and tourism industries and training in robot management. The implementation of hotel robots is usually integrated with other technologies such as facial recognition, automatic payment, drone delivery, and autonomous vehicles [9]. According to [39, 40], a robot control system requires three development levels, including hardware, functions, and services. Hardware refers to the shape of the mechanical design, which includes the main part of the service robot, the perception system (sensor), and the motion system (actuator). Function refers to the software architecture of the control system, which can implement navigation, dialogue, visual and speech recognition, positioning and mapping mechanisms, and the representation of knowledge mental models. Service refers to the added value that the hotel aims to create and provide customers with services to maintain a competitive advantage. Chung et al. [41] believe that a key requirement of the robot is that the

robot should perceive people around the robot. Through two field studies of custom programs and iterative design, a programming system for social interaction applications of mobile service robots has been developed. The timing for widespread deployment of service robots anywhere in the hotel will depend on when technical limitations can be addressed [42]. Zhang et al. [43] proposed a smart hotel robot based on ROS, which simplifies the check-in process. López et al. [44] proposed an automated hotel assistant system based on a series of mobile platforms that interact with guests to help them complete different tasks, including bring small items to customers, show them the different attractions of the hotel, accompany guests into their rooms, and provide them with general information.

*2.3. Related Work on Robot Service Quality and Feedback.* Customers can feel valuable in the self-service process when interacting with service robots out of fun, enjoyment, and curiosity [45, 46]. Robot customization with user-friendly interface design is important to attract specific new customer markets [47, 48]. Although current robotics technology may not directly complement comprehensive human services, from a marketing perspective, service robots will indirectly attract existing and potential customers [49, 50]. Zhang and Qi [51] pointed out that people with high education and high income who are interested in AI technology are more looking forward to staying in AI robot hotels. Some authors have even suggested that marketers should start thinking about robots as attractions [52]. With the popularity of robots, curiosity will no longer be an important reason for guests to stay in hotels but instead will be service quality and customer satisfaction. Many scholars have researched and explored the service quality of robots [53–55]. Service robots can improve perceived service quality through new attractive interactive services, communication, and contact with customers [56]. Unlike industrial robots whose performance indicators depend entirely on efficiency, the success of service robots depends on user satisfaction [57]. Robots that fail to deliver services can severely impact customer satisfaction and hotel performance [58]. Nakanishi et al. [59] found that the warm heart interactive service of humanoid robots potentially improves customer satisfaction with the entire service.

Customer satisfaction with robot services is largely determined by the robot itself. Therefore, the optimal design of the robot is particularly important. One of the current research studies is to subjectively obtain feedback through customers' questionnaires, polls [60], secondary data analysis [61], hybrid approach [62], and so on. Jeonghye et al. [14] proposed an evolutionary model of service robots with LCD touch panels, such as home robots, by questionnaire survey of parents and children and proposed the importance of rebuilding the robot's facial expressions based on the evaluation model. Tussyadiah and Park [9] found that customers' perception of different types of hotel service robots is different through online surveys and the use of biosensors to measure automatic emotional responses, and the adoption of hotel service robots is significantly affected

by the following aspects: intelligent and perceived security. Another aim is to objectively optimize the algorithm, for example, Ju et al. [18] designed a Python plug-in management system based on the mechanical mechanism design of the robot platform to implement graphical programming to control the robot. Huang et al. [63] proposed a multi-attention-based group recommendation model (MAGRM), and experiments showed that its performance in solving group recommendation problems is significantly better than the latest technology.

The above methods are either subjective or disconnected from the customer, and it is important to find an objective and customer-relevant robotic optimization feedback channel, such as robot optimization through service commands issued directly by customers. Therefore, this paper analyzes the general situation of hotel robot service from the command, by studying what the consumer's commands to the robot mainly include, whether it has formed a certain pattern according to the time series, and whether there are different service differences from the response of the command. The aim is to analyze the service content and quality of hotel robots and to point out new directions and ideas for hotel robot design. It is expected to contribute to the further optimization of robot design, such as packaging preset behavior patterns, optimizing repetitive commands, and human development to improve the service quality of the robot.

### 3. Materials and Methods

To have a systematic analysis on human-robot interaction, a rich data source containing a large amount of interaction records is needed. Then, the collected big dataset is pre-processed to remove incorrect and irrelevant information to reduce possible bias in results. After a basic statistical analysis on the clean dataset, the paper uses a frequent pattern mining method to study the user command patterns and a social network-based clustering method, which aims to help understand users' motivations, interests, and particular behaviors.

*3.1. Data Collection and Preprocessing.* The data were collected from 88 hotels in 23 cities across the mainland China. The service robots are equipped in 789 hotel rooms in these hotels. A robot management system by robot manufactory recorded all the user commands received by the hotel robots in the rooms. In this work, the historical data of command records in 2017 are selected for analysis and evaluation purpose due to the stability of the robot operations and the completeness of data.

In the interaction between user and hotel robots, the commands were given by voice. The user spoke out their requests and then the robot received the entire voice sentences and converted them into simple text-based command labels by its voice recognition functions. The collected data were then preprocessed to remove the inconsistent command labels caused by the updation of the robot system. The unrecognized commands were also removed as they were mainly caused by accent issues. After preprocessing, 745,528

valid user commands from 49,955 hotel customers were kept and input into the next-step analysis processing.

*3.2. Association Analysis on User Commands.* To understand human-robot interaction, the commands given by the users are the only resource that can be carefully studied in this case. Frequent pattern mining is used to discover most popular user command patterns in order.

*3.2.1. Sequential Frequent Pattern Discovery from User Commands.* Frequent patterns are often used to present the most common or similar features among data examples in a dataset. Given a set of commands, frequent pattern mining aims to find the rules that enable us to predict the occurrence of a specific item based on the occurrence of other items in the commands. A frequent pattern can be determined by a set of standard constraints [64].

The simple form for frequency of a pattern can be easily gained by counting the total number of its occurrence in the entire dataset. This is commonly used for finding out the most popular patterns with only one individual command.

Support, as one of the widely used constraints, was proposed by Agrawal et al. [65] in their Association Rule Mining (ARM) algorithm to identify a frequent pattern with two or more user commands. Let  $C$  be the complete set of all user commands in the collected dataset  $D$ .  $c_1$  and  $c_2$  be two commands from  $C$ , that is,  $c_1 \in C$ ,  $c_2 \in C$ .  $c_1$  and  $c_2$  are two different commands. The support of these two commands can be obtained by

$$\text{Supp}(c_1, c_2) = \frac{\text{number of interactions containing } c_1 \text{ and } c_2 \text{ in } D}{\text{total number of interactions in } D}. \quad (1)$$

A pattern can be indicated as a frequent one if its support score is not less than a minimum support threshold  $\delta$ ,  $\text{Supp}(c_1 \text{ and } c_2) \geq \delta$ . The classic Apriori-heuristic-based ARM is not suitable for processing large-scale dataset as it is very time consuming and expensive in computation. The concept of interestingness is then considered into the process to avoid generating too many useless candidate command patterns [66]. Only the candidate patterns with a higher interestingness score than the minimum interestingness threshold can join the next round of candidate pattern generation. A candidate pattern with two commands is generated based on the single commands with acceptable support score in previous generation; the candidate patterns with three commands are generated from the candidate patterns with two commands, and so on. The interestingness is calculated by the leverage of two commands as an example below:

$$|\text{leverage}(c_1, c_2)| = |\text{supp}(c_1, c_2) - \text{supp}(c_1)\text{supp}(c_2)| \geq \theta. \quad (2)$$

In such a way, the computation time and waste can be controlled under an acceptable range. In this work, the

candidate command patterns are indicated as frequent patterns based on their frequency of occurrence. One of the reasons is due to the high repeats of the same command patterns in one interaction. The user could place the same set of commands one after another, for example, “turn on lamp, turn off lamp, turn on lamp, play music, turn off lamp, turn on lamp ...” Therefore, the strength measurement constraints in ARM do not fit and are ignored in this study.

Another special feature in this work is that the actual time when users placed the commands was recorded as well, which makes the collected data time-series data. Time-series data are often presented in data sequences. In this work, each sequence records an interaction between the user and the robot. Accordingly, the command patterns are not random combination of the commands, and the order of the commands is also important.

*3.2.2. Social Network Analysis for User Clustering.* In recent years, the world has become more and more complex, and the research community has done a lot of work on the measurement of complexity. Social network is an analysis and simulation method that attempts to use the relationship between points and lines to analyze complex systems through the algorithmic characteristics of graphs. Social network analysis is widely used to understand the nature and discover useful patterns representing the relationships among members in a group [67]. In this work, the relationships among the hotel customers (users) were studied based on the commands they placed. The dataset was reorganized into a collection of the sequences of commands recorded for each individual user. A  $n \times n$  matrix  $M$  was obtained where  $n$  is the total number of the user commands. The intersection  $m_{ij}$  records the number of occasions that command  $c_i$  and command  $c_j$  are in the same clique. A clique is a subset of commands that always occur together over all the command sequences [68]. The total number of links in the network is calculated by  $l = \sum \sum m_{ij}$ .

Hierarchical clustering (Johnson, 1967) was then used to group commands into clusters based on the nearest pair of commands. By clustering the commands together, the users’ interests are discovered, that is, what commands are always called together with high densities. The social network analysis was implemented using a software called UCINET [69].

## 4. Results and Discussion

This section presents the experiment results obtained by applying the selected method on the collected data. The results are explained with the method used, respectively, followed by discussion of overall findings.

*4.1. Dataset and Descriptive Statistical Analysis.* After data cleaning and preprocessing, a total number of 103 individual commands were extracted from the collected historical data. Table 1 lists the top 50 commands placed

TABLE 1: Top 50 frequent individual user commands received by hotel service robots.

| Top # | Command              | Frequency |        |
|-------|----------------------|-----------|--------|
| 1     | OpenLamp             | 125,855   | 16.97% |
| 2     | CloseLamp            | 118,271   | 15.95% |
| 3     | TVOFF                | 103,759   | 13.99% |
| 4     | OpenCurtain          | 91,146    | 12.29% |
| 5     | CloseCurtain         | 78,049    | 10.53% |
| 6     | CloseScreen          | 41,324    | 5.57%  |
| 7     | MusicPlay            | 33,254    | 4.48%  |
| 8     | AirON                | 30,307    | 4.09%  |
| 9     | OpenDoor             | 17,435    | 2.35%  |
| 10    | TVON                 | 15,416    | 2.08%  |
| 11    | AirOFF               | 9,099     | 1.23%  |
| 12    | OpenPowerLight       | 8,879     | 1.20%  |
| 13    | MeetingModeON        | 2,842     | 0.38%  |
| 14    | TVChannelChange      | 2,629     | 0.35%  |
| 15    | TVVolUp              | 2,325     | 0.31%  |
| 16    | ServiceON            | 2,267     | 0.31%  |
| 17    | OpenBedroomLight     | 2,236     | 0.30%  |
| 18    | CloseWCLight         | 2,207     | 0.30%  |
| 19    | MusicStop            | 2,045     | 0.28%  |
| 20    | OpenWCLight          | 2,002     | 0.27%  |
| 21    | SleepModeON          | 1,974     | 0.27%  |
| 22    | CloseReadLight       | 1,972     | 0.27%  |
| 23    | CloseDengDaiLight    | 1,860     | 0.25%  |
| 24    | OpenNightLight       | 1,719     | 0.23%  |
| 25    | OpenLangLight        | 1,503     | 0.20%  |
| 26    | OpenDengDaiLight     | 1,390     | 0.19%  |
| 27    | OpenTopLight         | 1,378     | 0.19%  |
| 28    | MusicNext            | 1,376     | 0.19%  |
| 29    | OpenReadLight        | 1,291     | 0.17%  |
| 30    | OpenBedLight         | 1,265     | 0.17%  |
| 31    | BlueTooth            | 1,242     | 0.17%  |
| 32    | ClosePowerLight      | 1,206     | 0.16%  |
| 33    | OpenLivingroomLight  | 1,031     | 0.14%  |
| 34    | AirCondSet25         | 956       | 0.13%  |
| 35    | OpenJingQianLight    | 953       | 0.13%  |
| 36    | AirCool              | 934       | 0.13%  |
| 37    | AirCondSet24         | 933       | 0.13%  |
| 38    | CloseLangLight       | 921       | 0.12%  |
| 39    | OpenStudyLight       | 888       | 0.12%  |
| 40    | CloseLivingroomLight | 856       | 0.12%  |
| 41    | AirCondSet26         | 852       | 0.11%  |
| 42    | CloseTopLight        | 852       | 0.11%  |
| 43    | TVVolDown            | 851       | 0.11%  |
| 44    | AirCondSet20         | 769       | 0.10%  |
| 45    | CloseBedLight        | 759       | 0.10%  |
| 46    | AirHeat              | 750       | 0.10%  |
| 47    | MusicPrev            | 713       | 0.10%  |
| 48    | OpenBackgroundLight  | 713       | 0.10%  |
| 49    | OpenScreen           | 696       | 0.09%  |
| 50    | AirCondSet30         | 688       | 0.09%  |

by the users in the interaction with hotel robots. The top 12 commands have much higher frequency to be called by the users than the rest, which have over 90% of the listed commands. Close to 70% of user commands fall into one of the OpenLamp, CloseLamp, TVOFF, OpenCurtain, and CloseCurtain commands. That is, about 90 commands have less than 0.4% of chances to be called once in a year.

*4.2. Time-Based User-Robot Interaction Analysis.* Analyzing the active degrees of user commands helps to understand the peak and off-peak periods of human-robot interaction. An hour-based activity frequency chart is presented in Figure 1. Some common senses can be detected including the interactions which always start alone with users waking up in the morning from 6 am and increase to reach the first high point around 10 am. A small drop is found around 12 pm when users leave the rooms for lunch, and the number of commands climbs back quickly from 1 pm and remains stable until dinner time. The peak time is around 10 pm when users stay in the room to enjoy entertainment. It is a surprise to see that users still communicate with service robots after midnight. Finally, the quiet time starts from 2 am to 5 am. Many people stay late in the hotel rooms and talk to service robots.

A month-based activity frequency chart is presented in Figure 2. During the whole year, the service robots are busy in July and August and most of them are in idle or standby status in January and February. This matches the factor that during summer time, more hotel reservations and travels are made in general. The monthly-based interaction data also indicate that most of the users who lived in the hotel rooms equipped with service robots tried to interact with them almost every time. The service robots are accepted by users who stayed in the hotels anytime.

*4.3. User-Robot Interaction Pattern Discovery.* Frequent command patterns are expected to provide some knowledge about users' interests, behaviors, and the responsiveness performance of service robots.

*4.3.1. Command Patterns for Popular Command Combinations.* In human-robot interaction analysis, a main object is to find out the user interests when they request robots for services. The frequent command patterns can indicate what groups of commands users preferred to work with service robots. The frequent individual user commands are listed in Table 1. Tables 2–4 list top 10 frequent patterns as the permutations of two, three, and four commands, respectively. The pattern generation stopped at four-command level since the frequency and support score of the candidate five-command patterns were too low to have significant influence. The frequency was measured by the total number of occurrences including the repeats in the same interaction. The number of interactions when a pattern appeared at least once was also recorded. The difference between these two numbers indicate that many patterns have high repeats in a single interaction. The reasons to have this situation are complex and will be explained in the following context. Note that the order of commands is a key concept to be considered in all cases.

In the frequent two-command and three-command patterns, most of the patterns contain operations of turn on and off lamp (all lights), curtain, and TV. In the frequent four-command patterns, close screen (for windows) as a new command can be seen in most of the patterns. However, the number of commands increased in the

patterns, and the total number of occurrences dropped significantly. A common behavior from all the pattern sets that can be seen is that users like to turn off light, curtain, window screen, and TV together. The difference is just the order of turning which one first. For those patterns regarding TV, users turned off the TV more frequent than to turn it on because the TV was automatically turned on after the user entered or returned to the rooms in many hotels in China region. In addition, from all the frequent patterns in these three tables, the lamps (lights) and the curtains were turned on or off together. If the curtain was closed, then the lamp would be off too, and vice versa. It is not a common case to close the curtain and turn on the lights, from which one can claim that such operation was normally done before the user went to bed. It also confirmed that the human-robot interactions were active in the late hours of the day (as shown in Figure 1). The patterns of opening curtain and turning on the lamps may suggest the natural lighting in the rooms is insufficient.

A particular behavior is found from the analysis, in which the users tried to “play” with service robots by placing paired commands. That is, when the first command was given, immediately another command followed. The paired commands are always to turn on or off one device, such as lamps, TV, curtain, and air conditioner. This indicates that the users were bored and called those commands without a clear purpose. They might just want to talk to service robots for fun or kill time. Table 5 lists the top 20 paired commands in this category. Rather than turning on and off a particular light, users preferred more to turn all the lights on and off. For TV and air conditioner, the same situations are detected. Users did not give details in order but went for simple and short commands.

*4.3.2. Robot Responsiveness Analysis.* Although the human-robot interaction is identified as the indirect one that only one-way communication is applied, the responsiveness of the service robots to the user's commands is an interesting topic to be explore more in this section.

The responsiveness analysis carried out in this work is to look at how many times a user needs to repeat continuously the same command until the robot takes action or the user gives up. From the frequent pattern analysis above, there are some differences between the total number of the occurrences of the patterns and the number of the interactions having the patterns. They are not the same numbers, and in many cases, there are big gaps between them. The reason for this is that some commands are repeated many times in one interaction. The commands with repeats were closely observed, and the findings are summarized in Figure 3. The number of repeats for one single command daily is around 2.9, which indicates that the service robot's responsiveness still has a big room for improvement. Top 20 repeated commands were extracted from the analysis. By comparing with the initial count on top 50 frequent individual commands in Table 1, 17 out of 20 most frequent commands are found to have high repeats. Most of the commands were repeated within 2 seconds.

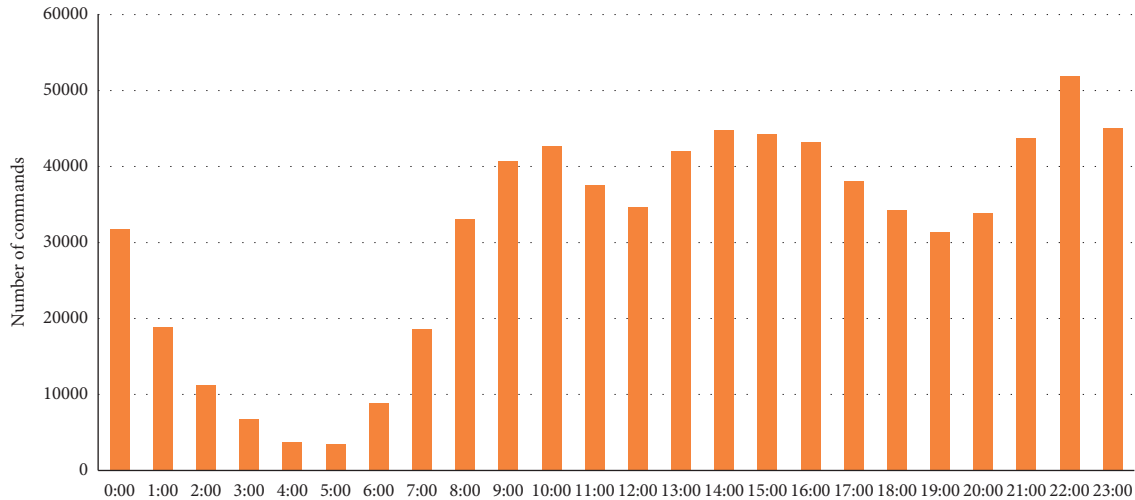


FIGURE 1: Hour-based human-robot interaction.

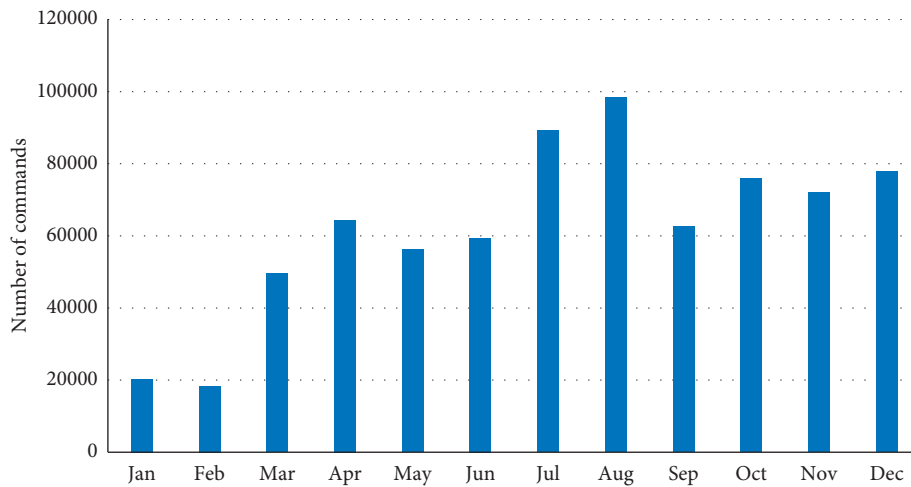


FIGURE 2: Monthly human-robot interaction.

TABLE 2: Top 10 frequent two-command patterns.

| Two-command frequent patterns | Frequency         |                     |
|-------------------------------|-------------------|---------------------|
|                               | Total occurrences | No. of interactions |
| CloseCurtain, CloseLamp       | 33,255            | 15,250 (30.53%)     |
| CloseScreen, TVOFF            | 31,976            | 14,284 (28.59%)     |
| CloseLamp, CloseScreen        | 28,713            | 13,092 (26.21%)     |
| OpenLamp, CloseLamp           | 21,253            | 12,973 (25.97%)     |
| TVOFF, CloseCurtain           | 16,870            | 9,245 (18.51%)      |
| OpenCurtain, CloseCurtain     | 11,531            | 7,859 (15.73%)      |
| OpenLamp, TVOFF               | 11,410            | 7,417 (14.85%)      |
| OpenLamp, OpenCurtain         | 10,641            | 8,467 (16.95%)      |
| TVOFF, OpenCurtain            | 9,554             | 6,692 (13.40%)      |
| CloseCurtain, OpenCurtain     | 8,253             | 5,946 (11.90%)      |

Table 6 lists the top 20 commands that have the most repeats. Not all of them have repeats due to the poor responsiveness of the robots. The repeats may be caused due to the nature of the commands themselves. For example, MusicVolDown, TVChannelADD, TVVolUp, MusicVolUp,

TVVolDown, and TVMenuUp are six commands that the users will repeat several times to reach their satisfied status. It is understandable that a user will keep on calling the command MusicVolDown to turn the music volume down to an acceptable amount. However, other commands to be repeated are treated as the cases of poor responsiveness of service robots. ReadLightMax has the highest repeat rate over 85% followed by OpenPowerLight (81.32%). These commands with high repeats need special attention to the service robot manufactory for performance.

**4.4. User Command Clustering.** The user clustering based on the commands they called in the human-robot interaction is shown in Figure 4. In total, six user clusters were obtained from social network analysis by UCINET. They are colored in Figure 4 for visualizing the user groups and the commands used by each group. Overall, a set of common commands is detected, which confirms the most popular commands like switching lights, pulling curtains, watching TV, and playing music are the favorites for all six groups of



TABLE 3: Top 10 frequent three-command patterns.

| Three-command frequent patterns      | Frequency         |                     |
|--------------------------------------|-------------------|---------------------|
|                                      | Total occurrences | No. of interactions |
| CloseCurtain, CloseLamp, CloseScreen | 26,062            | 12,141 (24.30%)     |
| CloseLamp, CloseScreen, TVOFF        | 26,421            | 12,227 (24.48%)     |
| CloseScreen, TVOFF, CloseCurtain     | 7,846             | 4,528 (9.03%)       |
| OpenLamp, CloseLamp, OpenLamp        | 6,547             | 4,512 (9.06%)       |
| CloseScreen, TVOFF, OpenLamp         | 6,128             | 4,356 (8.72%)       |
| CloseLamp, OpenLamp, CloseLamp       | 5,359             | 3,766 (7.54%)       |
| CloseCurtain, CloseScreen, TVOFF     | 4,896             | 3,596 (7.20%)       |
| OpenCurtain, OpenLamp, TVOFF         | 4,360             | 2,841 (5.69%)       |
| OpenCurtain, OpenLamp, CloseLamp     | 3,800             | 3,649 (7.30%)       |
| CloseScreen, TVOFF, OpenCurtain      | 3,608             | 2,724 (5.45%)       |

TABLE 4: Top 10 frequent four-command patterns.

| Four-command frequent patterns                    | Frequency         |                     |
|---|-------------------|---------------------|
|   | Total occurrences | No. of interactions |
| CloseCurtain, CloseLamp, CloseScreen, TVOFF       | 19,487            | 8,979 (17.97%)      |
| CloseLamp, CloseScreen, TVOFF, OpenLamp           | 4,974             | 3,558 (7.12%)       |
| CloseLamp, CloseScreen, TVOFF, OpenCurtain        | 2,708             | 2,009 (2.59%)       |
| OpenLamp, CloseCurtain, CloseLamp, CloseScreen    | 2,538             | 1,973 (3.59%)       |
| CloseLamp, OpenLamp, CloseLamp, OpenLamp          | 1,874             | 1,307 (2.62%)       |
| OpenLamp, CloseLamp, OpenLamp, CloseLamp          | 1,830             | 1,294 (2.59%)       |
| CloseLamp, CloseCurtain, CloseLamp, CloseScreen   | 1,578             | 1,438 (2.88%)       |
| MusicPlay, CloseCurtain, CloseLamp, CloseScreen   | 1,332             | 1,184 (2.37%)       |
| OpenCurtain, CloseCurtain, CloseLamp, CloseScreen | 1,206             | 1,018 (2.04%)       |
| OpenCurtain, OpenLamp, CloseLamp, OpenLamp        | 681               | 672 (1.35%)         |

TABLE 5: Top 20 paired commands.

| Command 1            | Command 2            | Frequency (total occurrences) |
|----------------------|----------------------|-------------------------------|
| OpenLamp             | CloseLamp            | 21,253                        |
| CloseLamp            | OpenLamp             | 19,884                        |
| OpenCurtain          | CloseCurtain         | 11,531                        |
| CloseCurtain         | OpenCurtain          | 8,253                         |
| TVOFF                | TVON                 | 7,832                         |
| TVON                 | TVOFF                | 4,162                         |
| AirON                | AirOFF               | 2,241                         |
| AirOFF               | AirON                | 1,524                         |
| MusicPlay            | MusicStop            | 420                           |
| MusicStop            | MusicPlay            | 253                           |
| OpenWCLight          | CloseWCLight         | 250                           |
| OpenLangLight        | CloseLangLight       | 222                           |
| CloseWCLight         | OpenWCLight          | 149                           |
| OpenTopLight         | CloseTopLight        | 124                           |
| CloseBackgroundLight | OpenBackgroundLight  | 118                           |
| OpenLivingroomLight  | CloseLivingroomLight | 117                           |
| OpenBedroomLight     | CloseBedroomLight    | 113                           |
| CloseLangLight       | OpenLangLight        | 105                           |
| OpenDengDaiLight     | CloseDengDaiLight    | 102                           |

users. Individual groups also have their preferred commands. Even with the same devices, they have different operations. Each group is discussed in more detail in the rest of the section.

Group 1 (in pink) has the biggest set of the commands. Except the commands related to switching TV, light, and curtain operations, this group works more active in playing

music and changing TV channels. The users in this group may prefer more to turn on TV to play music, but few of them used commands to adjust room temperature.

Group 2 (in dark blue) has the active users who used service robots to mainly adjust the room temperature by air conditioner. A set of temperature setting commands like “Aircondset25,” “Aircondset24,” “Aircondset26,”

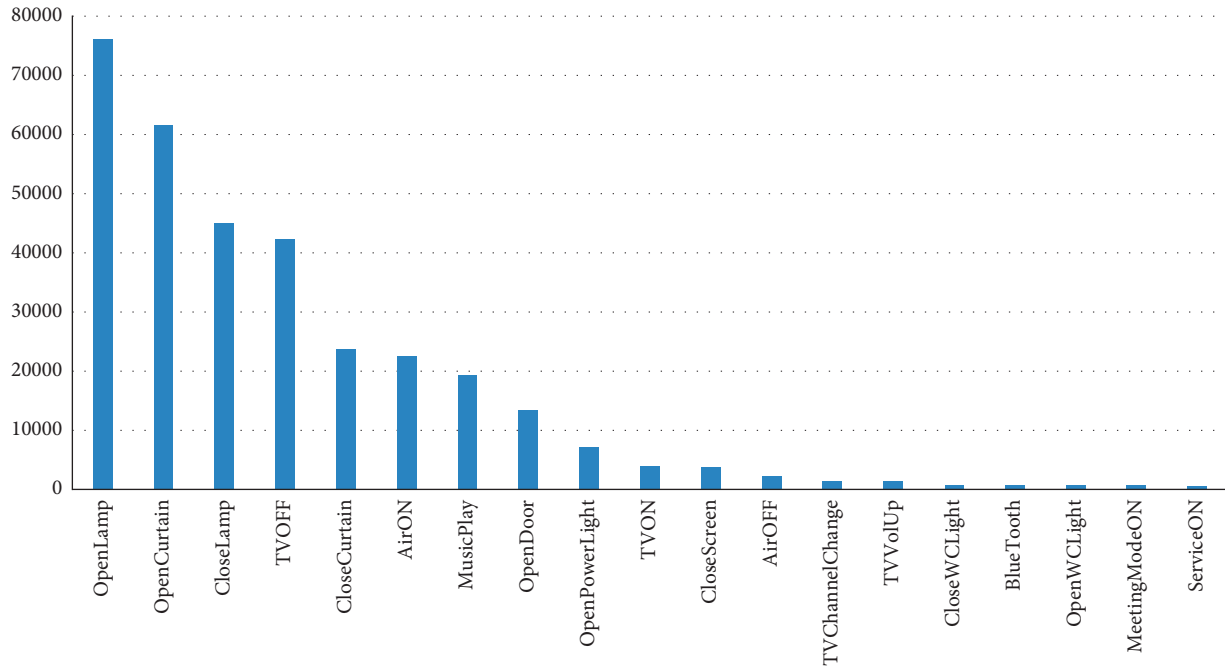


FIGURE 3: Top 20 repeated user commands.

TABLE 6: Top 20 commands with most repeats.

| Top # | Command              | Repeat rate (%) |
|-------|----------------------|-----------------|
| 1     | Read LightMAX        | 85.28           |
| 2     | OpenPowerLight       | 81.32           |
| 3     | OpenDoor             | 77.32           |
| 4     | AirON                | 74.49           |
| 5     | TVChannelSub         | 74.42           |
| 6     | MusicVolDown         | 67.94           |
| 7     | OpenCurtain          | 67.62           |
| 8     | TVChannelADD         | 64.42           |
| 9     | TVVolUp              | 60.82           |
| 10    | OpenLamp             | 60.50           |
| 11    | MusicPlay            | 57.89           |
| 12    | BlueTooth            | 57.57           |
| 13    | TVChannelChange      | 54.62           |
| 14    | MusicVolUp           | 54.03           |
| 15    | CloseBedLight        | 53.75           |
| 16    | OpenLivingroomLight  | 53.35           |
| 17    | TVVolDown            | 50.65           |
| 18    | CloseLivingroomLight | 49.30           |
| 19    | TVMenuUp             | 48.61           |
| 20    | CloseDoor            | 48.12           |

“Aircondset20,” and so on is detected in this group particularly. The temperature range is from 17°C to 29°C across the year, which falls into the normal external temperature adjustment of the human body. This user group focused more on the operations with accurate requirements rather than some general commands preferred by Group 1.

Group 3 (in light blue) is a small one at the margin areas. The commands used more by this group are entertainment-related operations on TV and playing music. They used “Bluetooth” to connect with their own smart devices and played music on TV. They also used “Checkoutn”

command to inform the service robot before they left the room. They can be indicated as avant-garde customers for hotels.

Group 4 (in green) often used robots to call for room service. The users in this group put “Nodisturbon” and “Cleanon” to call for cleaning service. They also watched the charge channels on TV rather than browsing the noncharge channels. These customers preferred more to keep the hotel room as their private space and had strong control on the activities in their space.

Group 5 (in yellow) and Group 6 (in brown) are two groups with the users who spent more time on switching lamps in the room to adjust the lighting. These two groups do not share a lot of overlaps with the other groups; especially Group 6 has commands with only one connection. These two groups have high interests to particular lights at different positions in the room. They did not use general commands to turn all the light on or off, but carefully and clearly gave the name of the lights. This behavior shows that the users moved a lot inside the room. They are energy savers to use individual lights in small areas.

*4.5. Discussion.* From the analysis on the results above, a number of issues and valuable points obtained need special attention from hotel management and technical experts in service robot manufactory.

In terms of theoretical contributions, this article makes an in-depth analysis of the services provided by robots from the perspective of the service industry. Different from the traditional robot research, it closely integrates human feelings and robot design and provides more knowledge to this area. Furthermore, this paper proposed a command mode for hotel robots through a complex network analysis

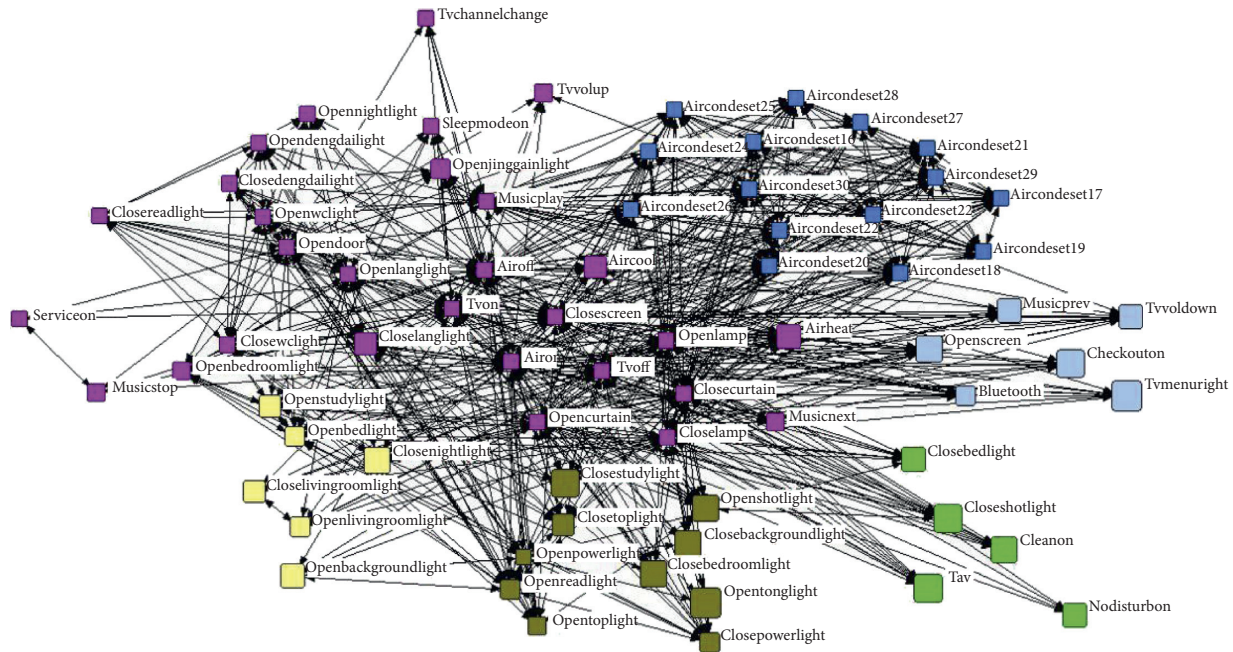


FIGURE 4: Six user clusters by social network analysis.

of a large number of robot commands. This is an innovative attempt in service robot research.

In terms of practical contributions, the robot needs to be improved in terms of human-computer interaction performance. From the same command in the same minute, the robot has certain problems in accepting customer orders, and there are many reasons for these problems. For example, the accent of the customer speaking Mandarin causes the robot not to be unrecognizable, or the voice of the customer is inaudible. Hotel robots lack semantic understanding of customer commands, which means that now hotel robots mostly use keywords matching and provide search results according to customers' command. Unrecognizable or too many external environmental interference factors cannot be removed by the robot. Regardless of the possible reasons, developers are still waiting to improve the performance of robots in human-computer interaction.

The customer conducts a curious operation. The typical performance is that the customer speaks many different types of operation commands in a short period of time, and thus the customer, being curious, is in a slow process of trying, trying, and accepting the service robot. Because of the curiosity about new things, the customer is willing to issue different commands to the robot. There are two kinds of mentality, one is the verification of the performance of the robot, and the other is the attempt of new things. Regardless of the mentality, it is conducive to the widespread use of future service robots. After repeated trials, the customer has a sensory feeling about the performance and service quality of the robot. In addition, because customers are more and more getting used to hotel robots and loss of interest, the paper provided a good reason for manufacturers to improve their service robots. This is crucial for the follow-up customer's willingness to continue using or staying in a hotel

with a robot. The following components can promote the customer's visit again; on the one hand, it is a good trial experience; on the other hand, it can bring convenience and comfort to customers. Of course, contrary to these two components, due to poor trial experience and complex or infinite operations, customers will be greatly reduced in the use and expectation of the robot.

Operation instructions are too limited. From the perspective of the type of operation instructions throughout the day, most of the operation instructions are control-type instructions. These instructions can only meet the customer's regular needs, and it is difficult to make a deep impression on the customer and urge the customer to have the urge to visit again. This is mainly due to the narrow range of hotel robot services. It is recommended that robot designers go to the hotel to experience the on-site experience of their products, think about the potential needs of customers from the perspective of customers, and develop some operational instructions that can impress customers. For example, automatic indoor temperature monitoring and adjustment, human body temperature automatic monitoring, potential hazard automatic alarm, customer demand input and implementation, and customer health care; only by developing these kinds of orders, we can grasp the customer's heart and promote and use service robots.

## 5. Conclusion

This paper proposed a first study on human-robot interaction between customers and service robots in hotels in China. Different from other service robot research studies that focused more on technical problems in robot design, development, and performance evaluation, this work made use of real data collected from live interactions between

customers and robots in the hotel rooms by applying frequent pattern mining and social network analysis techniques to find out hotel customers' interests, motivations, and behaviors when they interact with service robots.

Among a total of 103 user commands called by over 49,955 hotel customers during their stays in 88 hotels from 23 cities in China, the frequent command patterns indicate that most of the customers called service robots to do simple operations such as switching on or off the lights, TV, curtain, and window screen. Hotel customers have special preferences on services when they interacted with service robots. Some like to try a big range of service; some prefer to give accurate commands to adjust the room environment; some spend their time in hotel rooms for entertainment; some take time to enjoy themselves in hotel rooms as their private spaces; and others treat service robots as a company to kill time by repeating same set of commands continuously. Robots are also found to be slow to respond to some of the user commands, and potential improvement is expected to be made in near future.

Since this is the first work of this area, there were some limitations when the research was conducted. The raw data collected from the robot management system by manufacturer contain inconsistent command labels due to system updation. Only the data in 2017 are complete to be used in this work. The users' information is lacking in this work due to privacy concerns and since their personal information cannot be shared by hotels. In addition, whether social network analysis is applicable to robot commands needs to be carefully studied afterwards. However, these issues did not stop this work from discovering useful knowledge to understand hotel customers' interaction with service robots, and this is a remarkable step in human-robot interaction study in the future. The findings help hotel management to adjust the services that the customers are interested in and what the robots could provide.

### Data Availability

The data are owned by the robot company and could not be released.

### Conflicts of Interest

The authors declare that they have no known competing financial interest or personal relationships that could have appeared to influence the work reported in this paper.

### Authors' Contributions

Lina Zhong was responsible for conceptualization, methodology, writing, and funding acquisition. Liyu Yang contributed to literature review. Jia Rong reviewed and edited the manuscript and supervised the study. Xiaonan Li was responsible for software and formal analysis.

### Acknowledgments

This study was supported by the Beijing Social Science Foundation under grant nos. 19JDXCA005 and 18JDGLB013

and National Natural Science Foundation of China under grant no. 71673015.

### References

- [1] S. Ivanov and C. Webster, "Perceived appropriateness and intention to use service robots in tourism," in *International Conference in Nicosia*, pp. 237–248, Springer, Cham, Switzerland, 2019.
- [2] International Federation of Robotics, United Nations, and Economic Commission for Europe, *World Robotics, 2002: Statistics, Market Analysis, Forecasts, Case Studies and Profitability of Robot Investment*, United Nations Publications, New York, NY, USA, 2002.
- [3] T. Shibata, "An overview of human interactive robots for psychological enrichment," *Proceedings of the IEEE*, vol. 92, no. 11, pp. 1749–1758, 2004.
- [4] M. M. Wasko and S. Faraj, "Why should I share? Examining social capital and knowledge contribution in electronic networks of practice," *MIS Quarterly*, vol. 29, no. 1, pp. 35–57, 2005.
- [5] S. H. Ivanov and C. Webster, "Designing robot-friendly hospitality facilities," in *Proceedings of the Scientific Conference on Tourism. Innovations. Strategies*, pp. 74–81, Bourgas, Bulgaria, 2017.
- [6] M. Kim, S. Kim, S. Park, M.-T. Choi, M. Kim, and H. Gomiaa, "Service robot for the elderly," *IEEE Robotics & Automation Magazine*, vol. 16, no. 1, pp. 34–45, 2009.
- [7] N. K. Lakshmi, G. Rajakumar, D. N. M. Kumaran, I. Dinesh, V. Caleb, and H. Arshadh, "Design and fabrication of medicine delivery robots for hospitals," in *Proceedings of International Conference on Recent Trends in Computing, Communication & Networking Technologies (ICRTCCNT)*, Chennai, Tamilnadu, India, October 2019.
- [8] S. Ivanov, C. Webster, and A. Garenko, "Young Russian adults' attitudes towards the potential use of robots in hotels," *Technology in Society*, vol. 55, pp. 24–32, 2018.
- [9] I. P. Tussyadiah and S. Park, "Consumer evaluation of hotel service robots," in *Information and Communication Technologies in Tourism 2018*, pp. 308–320, Springer, Cham, Switzerland, 2018.
- [10] S. Thrun, "Toward a framework for human-robot interaction," *Human-Computer Interaction*, vol. 19, no. 1–2, pp. 9–24, 2004.
- [11] M. van Osch, D. Bera, K. van Hee, Y. Koks, and H. Zeegers, "Tele-operated service robots: rose," *Automation in Construction*, vol. 39, pp. 152–160, 2014.
- [12] S. Berezina, H. I. Stainslav, and C. Webster, "Adoption of robots and service automation by tourism and hospitality companies," *Revista Turismo & Desenvolvimento*, vol. 27, pp. 1501–1517, 2017.
- [13] A. Martins, "Brave new world: service robots in the frontline," *Journal of Service Management*, vol. 29, no. 5, pp. 907–931, 2018.
- [14] H. Jeonghye, L. Jaeyeon, and C. Youngjo, "Evolutionary role model and basic emotions of service robots originated from computers," in *Proceedings of the International Workshop on Robots and Human Interactive Communication*, pp. 30–35, IEEE, Nashville, TN, USA, August 2005.
- [15] D. Norman and S. W. Draper, *User Centered Design: New Perspectives on Human-Computer Interaction*, Lawrence Erlbaum Associates, Hillsdale, NJ, USA, 1986.
- [16] C. Webster and S. Ivanov, "Adoption of robots, artificial intelligence and service automation by travel, tourism and

- hospitality companies—a cost-benefit analysis,” in *Artificial Intelligence and Service Automation by Travel*, Emerald Group Publishing, Bingley, UK, 2017.
- [17] H. Jeonghye, J. Miheon, P. Sungju, and K. Sungho, “The educational use of home robots for children,” in *Proceedings of the International Workshop on Robots and Human Interactive Communication*, pp. 378–383, IEEE, Nashville, TN, USA, August 2005.
- [18] A. Ju, D. Chen, and J. Tang, “Design of Educational Robot Platform Based on Graphic Programming,” in *Advances in Intelligent Systems and Computing*, vol. 1117, Singapore: Springer, 2019. International Conference on Big Data Analytics for Cyber-Physical-Systems.
- [19] S. Lee and A. M. Naguib, “Toward a sociable and dependable elderly care robot: design, implementation and user study,” *Journal of Intelligent & Robotic Systems*, vol. 98, no. 11, pp. 5–17, 2019.
- [20] D. Bohus and E. Horvitz, “Directions robot: in-the-wild experiences and lessons learned,” in *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems*, pp. 637–644, Paris, France, May, 2014.
- [21] T. Kanda, M. Shiomi, Z. Miyashita, H. Ishiguro, and N. Hagita, “An affective guide robot in a shopping mall,” in *Proceedings of the International Conference on Human-Robot Interaction*, pp. 173–180, ACM/IEEE, San Diego, CA, US, October 2009.
- [22] T. Linder, S. Breuers, B. Leib, and K. O. Arras, “On multi-modal people tracking from mobile platforms in very crowded and dynamic environments,” in *Proceedings of the International Conference on Autonomous Robot Systems and Competitions*, pp. 5512–5519, IEEE, Stockholm, Sweden, May 2016.
- [23] H. S. Mirheydar and J. K. Parsons, “Diffusion of robotics into clinical practice in the United States: process, patient safety, learning curves, and the public health,” *World Journal of Urology*, vol. 31, no. 3, pp. 455–461, 2013.
- [24] J. Hu, Y. Sun, G. Li, G. Jiang, and B. Tao, “Probability analysis for grasp planning facing the field of medical robotics,” *Measurement*, vol. 141, pp. 227–234, 2019.
- [25] G. Ritzer, “Hospitality and prosumption,” *Research in Hospitality Management*, vol. 5, no. 1, pp. 9–17, 2015.
- [26] P. Alexis, “R-tourism: introducing the potential impact of robotics and service automation in tourism,” *Ovidius University Annals, Economic Sciences Series*, vol. 17, no. 1, pp. 211–216, 2017.
- [27] J. Crook, “Starwood introduces robotic butlers at aloft hotel in cupertino,” 2014, <https://techcrunch.com/2014/08/13/starwood-introduces-robotic-butlers-at-aloft-hotel-in-palo-alto/>.
- [28] Guardian, “Japan’s robot hotel: a dinosaur at reception, a machine for room service,” 2015, <https://www.theguardian.com/world/2015/jul/16/japans-robot-hotel-a-dinosaur-at-reception-a-machine-for-room-service>.
- [29] Hilton, “Hilton and IBM pilot ‘connie,’ the world’s first watson-enabled hotel concierge,” 2016, <http://newsroom.hilton.com/index.cfm/news/hilton-and-ibm-pilot-connie-the-worlds-first-watson-enabled-hotel-concierge>.
- [30] HotelManagement.net, “WYNN las vegas adds amazon echo to all guestrooms,” 2016, <http://www.hotelmanagement.net/tech/wynn-lasvegas-adds-amazon-echo-to-all-hotel-rooms>.
- [31] C. L. Liang, “Future hotel is coming, Alibaba opens the first unmanned hotel,” *Jing Rong Ke Ji Shi Dai*, vol. 11, p. 91, 2018.
- [32] C. Dirican, “The impacts of robotics, artificial intelligence on business and economics,” *Procedia—Social and Behavioral Sciences*, vol. 195, pp. 564–573, 2015.
- [33] C. B. Frey and M. A. Osborne, “The future of employment: how susceptible are jobs to computerisation?” *Technological Forecasting and Social Change*, vol. 114, pp. 254–280, 2017.
- [34] H. Osawa, A. Ema, H. Hattori et al., “What is real risk and benefit on work with robots? From the analysis of a robot hotel,” in *Proceedings of the Companion of the 2017 International Conference on human-robot interaction*, March 2017.
- [35] C.-M. Kuo, L.-C. Chen, and C.-Y. Tseng, “Investigating an innovative service with hospitality robots,” *International Journal of Contemporary Hospitality Management*, vol. 29, no. 5, pp. 1–17, 2017.
- [36] B. B. Stringam and J. H. Gerdes, “The automation of service in the hotel industry,” in *Proceedings of the Global Conference on Services Management (GLOSERV 2017)*, Volterra, Italy, October 2017.
- [37] J. Murphy, U. Gretzel, and J. Pesonen, “Marketing robot services in hospitality and tourism: the role of anthropomorphism,” *Journal of Travel & Tourism Marketing*, vol. 36, no. 7, pp. 784–795, 2019.
- [38] J. Murphy, C. Hofacker, and U. Gretzel, “Dawning of the age of robots in hospitality and tourism: challenges for teaching and research,” *European Journal of Tourism Research*, vol. 15, pp. 104–111, 2017.
- [39] T. Haidegger, M. Barreto, P. Gonçalves et al., “Applied ontologies and standards for service robots,” *Robotics and Autonomous Systems*, vol. 61, no. 11, pp. 1215–1223, 2013.
- [40] E. Zalama, J. G. García-Bermejo, S. Marcos et al., “Sacarino, a service robot in a hotel environment,” *ROBOT2013: First Iberian Robotics Conference, Advances in Intelligent Systems and Computing*, Springer, vol. 253, Cham, Switzerland, 2014.
- [41] M. J.-Y. Chung, J. Huang, L. Takayama, L. Takayama, T. Lau, and M. Cakmak, “Iterative design of a system for programming socially interactive service robots,” *Social Robotics*, Springer, vol. 9979, pp. 919–929, Cham, Switzerland, 2016.
- [42] Y.-H. Chin, H.-P. Lee, C.-W. Su et al., “A framework design for human-robot interaction,” in *Advanced Technologies, Embedded and Multimedia for Human-Centric Computing*, vol. 260, pp. 1043–1048, Springer, Lecture Notes in Electrical Engineering, Springer, 2014.
- [43] Y. Zhang, X. Wang, X. Wu, W. Zhang, M. Jiang, and M. Al-Khassawneh, “Intelligent hotel ROS-based service robot,” in *International Conference on Electro Information Technology (EIT)*, May 2019.
- [44] J. López, D. Pérez, E. Zalama, and J. Gómez-García-Bermejo, “BellBot—a hotel assistant system using mobile robots,” *International Journal of Advanced Robotic Systems*, vol. 10, no. 1, pp. 40–52, 2013.
- [45] M. S. Chtourou and N. Souiden, “Rethinking the TAM model: time to consider fun,” *Journal of Consumer Marketing*, vol. 27, no. 4, pp. 336–344, 2010.
- [46] C. Grönroos and A. Ravald, “Service as business logic: implications for value creation and marketing,” *Journal of Service Management*, vol. 22, no. 1, pp. 5–22, 2011.
- [47] C. Morosan and A. DeFranco, “Co-creation of value using hotel interactive technologies: examining intentions and conversion,” *International Journal of Contemporary Hospitality Management*, vol. 31, no. 3, pp. 1183–1204, 2019.
- [48] A. Scherer, N. V. Wunderlich, F. von Wangenheim, University of Paderborn, and E. T. H. Zürich, “The value of self-service: long-term effects of technology-based self-service usage on customer retention,” *MIS Quarterly*, vol. 39, no. 1, pp. 177–200, 2015.
- [49] M. Barrett, University of Cambridge, E. Davidson et al., “Service innovation in the digital age: key contributions and future directions,” *MIS Quarterly*, vol. 39, no. 1, pp. 135–154, 2015.

- [50] D. Kindström, C. Kowalkowski, and E. Sandberg, "Enabling service innovation: a dynamic capabilities approach," *Journal of Business Research*, vol. 66, no. 8, pp. 1063–1073, 2013.
- [51] Y. Zhang and S. Qi, "User experience study: the service expectation of hotel guests to the utilization of ai-based service robot in full-service hotels," in *HCI in Business, Government and Organizations. eCommerce and Consumer Behavior*, F. H. Nah and K. Siau, Eds., vol. 11588 Cham, Switzerland-Springer, 2019 Lecture Notes in Computer Science International Conference on Human-Computer Interaction.
- [52] S. H. Ivanov and C. Webster, "The robot as a consumer: a research agenda," in *Proceedings of the Marketing: Experience and Perspectives Conference*, University of Economics-Varna, Bulgaria, pp. 71–79, June 2017.
- [53] Y. Choi, M. Choi, M. Oh, and S. Kim, "Service robots in hotels: understanding the service quality perceptions of human-robot interaction," *Journal of Hospitality Marketing & Management*, pp. 1–23, 2019.
- [54] Y. Kim and H. S. Lee, "Quality, perceived usefulness, user satisfaction, and intention to use: an empirical study of ubiquitous personal robot service," *Asian Social Science*, vol. 10, no. 11, pp. 1–16, 2014.
- [55] R. Pinillos, S. Marcos, R. Feliz, E. Zalama, and J. Gómez-García-Bermejo, "Long-term assessment of a service robot in a hotel environment," *Robotics and Autonomous Systems*, vol. 79, pp. 40–57, 2016.
- [56] C.-M. Kuo, L.-C. Chen, and C.-Y. Tseng, "Investigating an innovative service with hospitality robots," *International Journal of Contemporary Hospitality Management*, vol. 29, no. 5, pp. 1305–1321, 2017.
- [57] C. Bartneck, T. Kanda, O. Mubin, and A. Al Mahmud, "Does the design of a robot influence its animacy and perceived intelligence?" *International Journal of Social Robotics*, vol. 1, no. 2, pp. 195–204, 2009.
- [58] S. Melián-González and J. Bulchand-Gidumal, "A model that connects information technology and hotel performance," *Tourism Management*, vol. 53, pp. 30–37, 2016.
- [59] J. Nakanishi, I. Kuramoto, J. Baba, O. Kohei, Y. Yoshikawa, and H. Ishiguro, "Can a humanoid robot engage in heart-warming interaction service at a hotel?" in *Proceedings of the 6th International Conference on Human-Agent Interaction*, Southampton, UK, December 2018.
- [60] J. Hudson, M. Orviska, and J. Hunady, "People's attitudes to robots in caring for the elderly," *International Journal of Social Robotics*, vol. 9, no. 2, pp. 199–210, 2017.
- [61] H. Yan, M. H. Ang, and A. N. Poo, "A survey on perception methods for human-robot interaction in social robots," *International Journal of Social Robotics*, vol. 6, no. 1, pp. 85–119, 2014.
- [62] M. Pino, M. Boulay, F. Jouen, and A. S. Rigaud, "Are we ready for robots that care for us?" attitudes and opinions of older adults toward socially assistive robots," *Frontiers in Aging Neuroscience*, vol. 7, p. 141, 2015.
- [63] Z. Huang, X. Xu, H. Zhu, and M. Zhou, "An efficient group recommendation model with multiattention-based neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, 2020, in Press.
- [64] X. Wu, C. Zhang, and S. Zhang, "Efficient mining of both positive and negative association rules," *ACM Transactions on Information Systems (TOIS)*, vol. 22, no. 3, pp. 381–405, 2004.
- [65] R. Agrawal, T. Imieinski, and A. Swami, "Mining association rules between sets of items in large databases," in *Proceedings of the ACM SIGMOD International Conference on the Management of Data*, P. Buneman and S. Jajodia, Eds., pp. 207–216, Washington, DC, USA, May 1993.
- [66] G. Li, R. Law, J. Rong, and H. Q. Vu, "Incorporating both positive and negative association rules into the analysis of outbound tourism in Hong Kong," *Journal of Travel & Tourism Marketing*, vol. 27, no. 8, pp. 812–828, 2010.
- [67] C. Hu and P. Racherla, "Visual representation of knowledge networks: a social network analysis of hospitality research domain," *International Journal of Hospitality Management*, vol. 27, no. 2, pp. 302–312, 2008.
- [68] S. P. Borgatti, M. G. Everett, and J. C. Johnson, *Analyzing Social Networks*, Sage Publications Ltd., Thousand Oaks, CA, USA, 2013.
- [69] S. P. Borgatti, M. G. Everett, and L. C. Freeman, *UCINET for Windows: Software for Social Network Analysis*, Analytic Technologies, Lexington, KY, USA, 2002.

## Research Article

# Decision Optimization of Low-Carbon Dual-Channel Supply Chain of Auto Parts Based on Smart City Architecture

Zheng Liu <sup>1</sup>, Bin Hu <sup>1</sup>, Bangtong Huang <sup>1</sup>, Lingling Lang <sup>1</sup>, Hangxin Guo <sup>1</sup>,  
and Yuanjun Zhao <sup>2</sup>

<sup>1</sup>School of Management, Shanghai University of Engineering Science, Shanghai 201620, China

<sup>2</sup>School of Business Administration, Shanghai Lixin University of Accounting and Finance, Shanghai 201209, China

Correspondence should be addressed to Yuanjun Zhao; 1139089@mail.dhu.edu.cn

Received 27 March 2020; Revised 24 April 2020; Accepted 5 May 2020; Published 21 May 2020

Guest Editor: Zhihan Lv

Copyright © 2020 Zheng Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Affected by the Internet, computer, information technology, etc., building a smart city has become a key task of socialist construction work. The smart city has always regarded green and low-carbon development as one of the goals, and the carbon emissions of the auto parts industry cannot be ignored, so we should carry out energy conservation and emission reduction. With the rapid development of the domestic auto parts industry, the number of car ownership has increased dramatically, producing more and more CO<sub>2</sub> and waste. Facing the pressure of resources, energy, and environment, the effective and circular operation of the auto parts supply chain under the low-carbon transformation is not only a great challenge, but also a development opportunity. Under the background of carbon emission, this paper establishes a decision-making optimization model of the low-carbon supply chain of auto parts based on carbon emission responsibility sharing and resource sharing. This paper analyzes the optimal decision-making behavior and interaction of suppliers, producers, physical retailers, online retailers, demand markets, and recyclers in the auto parts industry, constructs the economic and environmental objective functions of low-carbon supply chain management, applies variational inequality to analyze the optimal conditions of the whole low-carbon supply chain system, and finally carries out simulation calculation. The research shows that the upstream and downstream auto parts enterprises based on low-carbon competition and cooperation can effectively manage the carbon footprint of the whole supply chain through the sharing of responsibilities and resources among enterprises, so as to reduce the overall carbon emissions of the supply chain system.

## 1. Introduction

Smart city is a new concept and mode to promote intelligent urban planning, construction, management, and service by using the new generation of information technology such as the Internet of Things, cloud computing, big data, and spatial geographic information integration. The development of a smart city in China attaches great importance to the integration of the smart city and green low carbon [1]. With the global warming, glacier melting, and other environmental problems becoming more and more serious, the auto parts enterprises should not ignore the importance of environmental protection while pursuing economic interests. We should actively coordinate the relationship between human

beings and nature, so as to achieve sustainable development. The Kyoto protocol adopted in 1997 has set clear emission reduction targets for the countries that signed it. At the same time, it clearly takes the market mechanism as a new path to solve the problem of greenhouse gas emission reduction, which promotes the symbiotic model of carbon emission responsibility sharing and resource sharing. China's demand for establishing a carbon trading mechanism is very urgent; on the one hand, in order to achieve the dual goals of low cost and low emissions and to maintain the long-term economic vitality of enterprises, and on the other hand, to strive for the carbon pricing right and to master the initiative of resource allocation in the low-carbon economy [2]. Therefore, with the steady development of the domestic

carbon emission trading market and the rapid growth of trading volume, the research on the optimization of low-carbon supply chain decision-making based on carbon emission responsibility sharing and resource sharing is of great significance for promoting energy conservation and emission reduction and promoting the transformation of the economic development mode and upgrading of the industrial structure.

Based on the current situation of the construction of a traditional low-carbon city, Ma et al. analyzed the existing problems and the advantages of an intelligent city and then put forward the countermeasures to promote the construction of an intelligent city and the specific path to solve all kinds of problems in the construction of an intelligent low-carbon city [3]. Based on the low-carbon value chain, Yin and others made in-depth exploration on the main components of the low-carbon value chain in the automobile industry and built a three-dimensional low-carbon path system of the automobile supply chain [4]. Lee used the case study method to improve our understanding of the carbon footprint of the automobile supply chain under the management background, providing a new way for the integration of carbon emissions in the supply chain management [5]. Yang and Guitao took the minimization of carbon emissions and the maximization of enterprise profits as the dual objectives and comprehensively used the dual theory and variational inequality to obtain the equilibrium conditions. The research shows that the environmental protection performance and economic benefit of enterprises are improved when the proportion of environmental protection target weight increases [6]. Wei considered the situation of carbon trading, introduced the change of carbon emissions in different forms of research and development, and constructed a differential game model to get the optimal research and development strategy of supply chain emission reduction [7]. Wang and Han established the loss function of perishable products with the goal of reducing carbon emission and total cost and studied the supply chain optimization strategy of location path inventory combination [8]. Díaz-Trujillo et al. analyzed the carbon pricing tools of the energy sector in Mexico and found that the carbon tax and the carbon trading system are more conducive to improve economic and environmental benefits than the traditional carbon pricing policy by using the multiobjective optimization formula [9]. Efthymiou and Papatheodorou established a low-cost and low-carbon multiobjective decision-making model and verified that cost reduction and carbon emission reduction can meet the requirements at the same time [10]. Battini et al. established a two-objective model related to cost and carbon emission and analyzed the model according to the change of carbon price, which showed that the low carbon price could not stimulate the sustainable purchase behavior [11].

Zhang et al. took low carbon emissions as the goal, considered the constraints of customer demand time window and loading and unloading time, and used the hybrid particle swarm optimization algorithm to optimize the multiproduct and multiobjective supply chain [12]. Lanzirui et al. built a minimum cost model under the background of

low carbon and analyzed the influencing factors of production decisions of enterprises with high carbon emissions under the background of carbon emission quota [13]. Taleizadeh et al. used two scenarios to study the interaction between carbon emission reduction, return policies, and quality improvement efforts. The research showed that higher refund prices are beneficial to reduce carbon emissions and improve product quality and supply chain profits [14]. Noura et al. studied the impact of carbon emission sensitive demand on supply chain design decision [15]. Huang et al. proposed a new group recommendation model based on multiattention, which uses the deep neural network structure based on multiattention to achieve accurate group recommendation [16]. Purohit et al. studied the inventory lot size problem based on the constraints of carbon emission and cyclic service level. The research showed that cost, inventory, and carbon emission decrease with the increase in carbon price [17]. Saxena et al. constructed a supply chain design model that comprehensively considered carbon tax policy, economy, and carbon emission and used the improved cross entropy method to get the carbon price that can obtain the maximum environmental benefit [18].

Mauro et al. evaluated the pretreatment technology of a coal-fired power plant from two aspects of environment and technology economy and evaluated the impact of pretreatment technology on the carbon emission avoidance cost [19]. Di Filippo et al. compared and analyzed the effects of carbon pricing policy, command and control policy, and resource incentive three tools in order to solve the obstacles of market failure to take measures to reduce carbon emissions in the concrete supply chain. The research showed that the carbon pricing policy is relatively effective [20]. Aiming at the problem of grab point planning, Hu et al. proposed a grab planning method based on the Gaussian process classification of big data [21]. On calculating the carbon footprint of the grain, Roibás et al. proposed alternative transportation schemes that can reduce the carbon footprint of final products [22]. Wilson and Staffell found that the transformation of coal to natural gas can be realized through effective carbon pricing, so as to reduce carbon emissions [23]. Allevi et al. established a model including carbon emission policy, recovery, transportation, and technical factors to evaluate the effect of applying environmental policies on a multilayer closed-loop supply chain network and solved it by using the theory of variational inequality. The analysis shows that the combined application of the carbon emission tax of truck transportation and the EU emission trading system (EU-ETS) at manufacturer level is beneficial to environmental protection [24]. In order to improve the effectiveness of Internet of Things recommendation, Huang et al. proposed a multimodal representation learning-based Internet of Things recommendation model [25]. Shaw et al. used the Benders decomposition algorithm to solve the problem of the sustainable supply chain network design under the constraints of opportunities. By using an example, we found that the carbon credit price is positively correlated with the enterprise's changeable cost and negatively correlated with the variable emissions [26]. Yu et al. established a mixed integer nonlinear programming



(MINLP) model of shelf space allocation under the background of the carbon tax and discussed the influence of shelf space capacity on product allocation decision [27]. Singh et al. considered the carbon footprint in the traditional supplier selection process and proposed a framework to help reduce the carbon footprint of beef products using the combination of big data, operational research, and other technologies [28].

To sum up, the current research related to the carbon footprint and the low-carbon supply chain focuses on the optimization of the supply chain system of carbon emission, while the research on carbon emission responsibility sharing and resource sharing is less. Therefore, in the context of carbon emission, this paper establishes a low-carbon supply chain decision-making optimization model based on carbon emission responsibility sharing and resource sharing. This model takes economy and environment as the goal and uses variational inequality to analyze the optimal conditions of the whole low-carbon supply chain system by coordinating and transforming the two interactive goal equilibrium problems. On the basis of proving the existence of the solution of the variational inequality of the supply chain system, this paper analyzes the change rule of the profit and carbon emission under the carbon trading with various parameters (such as recycling coefficient and remanufacturing coefficient of waste products) by numerical simulation and verifies the rationality and effectiveness of the model. It can be seen from the results that through the competition and cooperation of upstream and downstream auto parts enterprises for low carbon, the carbon footprint of the whole supply chain can be effectively managed, so as to achieve the purpose of reducing the overall carbon emissions of the supply chain system.

The innovations of this paper are as follows: (1) from the perspective of research, in addition to considering carbon emission constraints, carbon emission responsibility sharing and resource sharing are also considered. (2) In the research method, the variational inequality is used to optimize the whole low-carbon supply chain system. (3) As for the research results, this paper draws the conclusion through simulation: through the competition and cooperation of upstream and downstream auto parts enterprises for low carbon, the carbon footprint of the whole supply chain can be effectively managed, so as to achieve the purpose of reducing the overall carbon emissions of the supply chain system.

## 2. Problem Description and Symbol Definition

The low-carbon supply chain network of auto parts is composed of suppliers, producers, physical retailers, online retailers, demand markets, and recyclers. The basic structure is shown in Figure 1. The solid line in Figure 1 represents the forward logistics, and the dotted line represents the reverse logistics. Auto parts raw material suppliers provide raw materials for manufacturers. The products of manufacturers are sold to the demand market through physical retailers and online retailers, and the products are recycled by recyclers

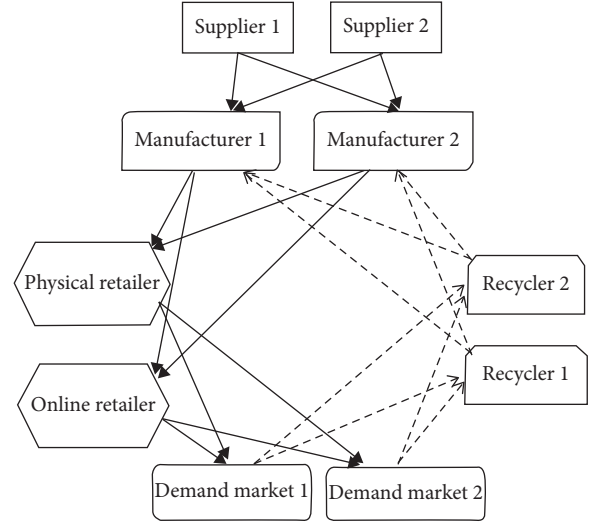


FIGURE 1: Network structure of the low-carbon supply chain.

after use. The definition of variables and function setting is shown in Tables 1 and 2, respectively.

## 3. The Establishment of Network Equilibrium Optimization Model for Supply Chain of Auto Parts

### 3.1. The Establishment of Network Equilibrium Model under the Economic Goal of Low-Carbon Supply Chain

**3.1.1. Network Equilibrium Model of Suppliers.** If  $Q_{sm}$  is the supply quantity from the supplier to the manufacturer,  $\rho_{sm}$  is the price of the material from the supplier to the manufacturer, and the profit of the supplier is the profit of the material supplied to the manufacturer minus the purchase cost of the supplier, then minus the transaction variable cost with the manufacturer, the objective function of the supplier  $s$  is

$$\max \sum_{m=1}^M \rho_{sm} Q_{sm} - \left( \sum_{m=1}^M f_s(Q_{sm}) + \sum_{m=1}^M c_{sm}(Q_{sm}) \right). \quad (1)$$

The optimal conditions of all suppliers are equivalent to the following variational inequalities, and the solution  $Q_{sm} \geq 0$  satisfies

$$\sum_{s=1}^S \sum_{m=1}^M \left[ \frac{\partial f_s(Q_{sm})}{\partial Q_{sm}} + \frac{\partial c_{sm}(Q_{sm})}{\partial Q_{sm}} - \rho_{sm} \right] \times (Q_{sm} - Q_{sm}^*) \geq 0, \quad \forall Q_{sm} \geq 0. \quad (2)$$

The variational inequality (2) shows that the transaction price of auto parts suppliers is equal to the sum of the marginal purchase cost and the marginal transaction cost for any supply greater than zero in the equilibrium state.

**3.1.2. Network Equilibrium Model of Manufacturers.** If  $Q_{mr}$  is the product quantity provided by the manufacturer to the

TABLE 1: Decision variables and parameters.

| Variable or parameter  | Definition of variable or parameter   |
|--|---|
| $s$  | Auto parts supplier, $s \in \{1, 2, \dots, S\}$   |
| $m$  | Auto parts manufacturer, $m \in \{1, 2, \dots, M\}$   |
| $r$  | Physical retailer and online retailer of auto parts, $r \in \{r_1, r_2\}$   |
|  | Physical retailer $r_1 \in \{1, 2, \dots, R\}$ ; online retailer $r_2 \in \{1, 2, \dots, R\}$   |
| $d$  | Auto parts market, $d \in \{1, 2, \dots, D\}$   |
| $n$  | Auto parts recycler, $n \in \{1, 2, \dots, N\}$   |
| $Q_{sm}, Q_{mr}, Q_{r,d}, Q_{dn}, Q_{mn}$  | The trading volume of auto parts supplier $s$ and manufacturer $m$ , auto parts manufacturer $m$ and physical retailer and online retailer $r$ , auto parts physical retailer and online retailer $r$ and demand market $d$ , auto parts demand market $d$ and recycler $n$ , auto parts recycler $n$ and manufacturer $m$  |
| $\rho_{sm}, \rho_{mr}, \rho_{r,d}, \rho_{dn}, \rho_{mn}, \rho_d$                         | Transaction price or sales price of auto parts supplier, manufacturer, physical retailer, online retailer, demand market, and recycler  |
| $f_s, f_{ms}, f_{mn}$  | Purchasing cost function of auto parts suppliers, production cost function of new materials used by manufacturers, and production cost function of old materials used   |
| $c_{sm}, c_{ms}, c_{mr}, c_{nr}, c_r, c_m, c_{r,d}, c_{dr}, c_{dn}, c_n, c_{nd}, c_{mn}$ | Transaction cost function of auto parts suppliers and manufacturers; transaction cost function of auto parts manufacturers and suppliers, physical retailers and online retailers, recyclers; storage cost function of auto parts physical retailers and online retailers, transaction cost function of physical retailers, online retailers and producers, transaction cost function of demand market; transaction cost function of auto parts demand market, physical retailers, online retailers and recyclers; auto parts recycling cost function of acquisition, transportation and storage, recycler and demand market, and transaction cost function of producer |
| $v_{im}, v_r, v_{in}$  | Production, sales, and recovery of auto parts and components unit energy use cost of each process input   |
| $u_{om}, u_{or}, u_{on}$   | Production, sales, and recovery of auto parts unit carbon emission cost of each process output end  |
| $CC_j^+, CC_j^-$   | The amount of carbon emissions purchased and sold by a node auto parts enterprise $j$ is calculated in tons of carbon dioxide equivalent ( $tCO_2e$ ), $j \in MURUN$  |
| $A^c, V^{cc}$  | Unit purchase and sale price of carbon trading  |
| $CF_{im}, CF_{ir}, CF_{in}$  | Weight coefficient of energy use at input nodes of production, sales, and recovery processes of auto parts  |
| $EF_{om}, EF_{or}, EF_{on}$  | Carbon emission coefficient of output nodes of production, sales, and recovery processes of auto parts  |
| $\phi_{sm}, \phi_{mn}$   | Availability factor of raw materials from auto parts suppliers or reprocessed products from recyclers   |
| $D_d(\rho_d)$  | Demand function of auto parts market $d$  |
| $\theta$   | Conversion and reuse coefficient of recycled auto parts   |
| $\chi$   | Disposal cost of unit scrapped auto parts   |
| $CO_j^{in}, CO_0^{out}$  | Characteristic coefficient of energy consumption and carbon emission converted into carbon dioxide equivalent   |
| $L^{COP}, L^{COS}, L^{CO_2e}$  | The maximum amount of carbon emission that can be purchased or sold and the amount of carbon emission distribution  |

TABLE 2: Function settings.

| Cost  | Cost function   |
|---|---|
| Purchase cost of auto parts supplier  | $f_1(Q_{sm}) = (\sum_{m=1}^2 Q_{1m})^2 + a(\sum_{m=1}^2 Q_{1m})(\sum_{m=1}^2 Q_{2m}) + b\sum_{m=1}^2 Q_{2m}$  |
| Transaction costs between auto parts suppliers and manufacturers                        | $f_2(Q_{sm}) = c(\sum_{m=1}^2 Q_{2m})^2 + (\sum_{m=1}^2 Q_{1m})(\sum_{m=1}^2 Q_{2m}) + \sum_{m=1}^2 Q_{1m}$<br>$c_{sm}(Q_{sm}) = aQ_{sm}^2 + eQ_{sm}$ , where $s = 1, 2; m = 1, 2$  |
| Production cost of new materials used by auto parts manufacturers                       | $f_{1s}(Q_{sm}, \varphi_{sm}) = (\varphi_{sm} \sum_{s=1}^2 Q_{s1})^2 + b(\varphi_{sm} \sum_{s=1}^2 Q_{s1})(\varphi_{sm} \sum_{s=1}^2 Q_{s2}) + f$   |
| Production cost of old materials used by auto parts manufacturers                       | $f_{2s}(Q_{sm}, \varphi_{sm}) = (\varphi_{sm} \sum_{s=1}^2 Q_{s2})^2 + (\varphi_{sm} \sum_{s=1}^2 Q_{s1})(\varphi_{sm} \sum_{s=1}^2 Q_{s2}) + g$<br>$f_{1n}(Q_{nm}, \varphi_{nm}) = a(\varphi_{nm} \sum_{n=1}^2 Q_{n1})^2 + b(\varphi_{nm} \sum_{n=1}^2 Q_{n1}) + b$<br>$f_{2n}(Q_{nm}, \varphi_{nm}) = (\varphi_{nm} \sum_{n=1}^2 Q_{n2})^2 + b(\varphi_{nm} \sum_{n=1}^2 Q_{n2}) + h$ |
| Transaction costs between auto parts manufacturers and suppliers                        | $c_{ms}(Q_{sm}) = aQ_{sm}^2 + kQ_{sm}$ , where $m = 1, 2; s = 1, 2$   |
| Transaction costs of auto parts manufacturers, physical retailers, and online retailers | $c_{nr}(Q_{nr}) = aQ_{nr}^2 + kQ_{nr}$ , where $m = 1, 2; r = 1, 2$   |
| Transaction costs of auto parts manufacturers and recyclers                             | $c_{nm}(Q_{nm}) = aQ_{nm}^2 + kQ_{nm}$ , where $m = 1, 2; n = 1, 2$   |
| Storage cost of auto parts physical retailer and online retailer                        | $c_1(Q_{m1}) = a(\sum_{m=1}^2 Q_{m1})^2 + k$ ,<br>$c_2(Q_{m2}) = a(\sum_{m=1}^2 Q_{m2})^2 + f$  |
| Transaction cost of auto parts physical retailer, online retailer, and manufacturer     | $c_{nr}(Q_{nr}) = aQ_{nr}^2 + lQ_{nr}$ , where $r = 1, 2; m = 1, 2$   |
| Transaction cost of auto parts physical retailer, online retailer, and demand market    | $c_{rd}(Q_{rd}) = aQ_{rd}^2 + kQ_{rd}$ , where $r = 1, 2; d = 1, 2$   |
| Auto parts demand function  | $D_1(\rho_d) = -b\rho_1 - o\rho_2 + p$<br>$D_2(\rho_d) = -o\rho_1 - b\rho_2 + p$  |
| Auto parts demand market and retailer's transaction cost                                | $c_{dr}(Q_{rd}) = Q_{rd} + f$ , where $d = 1, 2; r = 1, 2$  |
| Auto parts demand market and transaction cost of recycler                               | $c_{dn}(Q_{dn}) = Q_{dn} + h$ , where $d = 1, 2; n = 1, 2$  |
| Storage cost of auto parts recycler   | $c_1(Q_{d1}) = a(\sum_{d=1}^2 Q_{d1})^2 + h$ ,<br>$c_2(Q_{d2}) = a(\sum_{d=1}^2 Q_{d2})^2 + b$  |
| Transaction cost between auto parts recycler and demand market                          | $c_{nd}(Q_{nd}) = qQ_{dn}^2 + Q_{dn}$ , where $n = 1, 2; d = 1, 2$  |
| Transaction cost between auto parts recycler and manufacturer                           | $c_{nm}(Q_{nm}) = aQ_{nm}^2 + Q_{nm}$ , where $n = 1, 2; m = 1, 2$  |
| Conversion and reuse coefficient of recycled auto parts                                 | $\theta = cr_k - t$   |
| Availability factor of reprocessed products of auto parts recycler                      | $\varphi_{sm} = \alpha\mu - u$  |

physical retailer and the online retailer,  $\rho_{mr}$  is the product price of the manufacturer, the cost of the manufacturer includes the production cost, the cost paid to the supplier and the recycler, the transaction cost with the supplier, the physical retailer, the online retailer, and the recycler, and the

cost of energy and carbon emission consumed in the production shall also be considered, then assuming that the unit energy cost consumed in the production process is  $v_i$  and the unit carbon emission cost is  $u_o$ , the objective function of producer  $m$  is

$$\begin{aligned}
\max \quad & \sum_{r=1}^R \rho_{mr} Q_{mr} - \left( \sum_{s=1}^S f_{ms}(Q_{sm}, \varphi_{sm}) + \sum_{n=1}^N f_{mn}(Q_{nm}, \varphi_{nm}) \right. \\
& + \sum_{s=1}^S \rho_{sm} Q_{sm} + \sum_{n=1}^N \rho_{nm} Q_{nm} + \sum_{s=1}^S c_{ms}(Q_{sm}) + \sum_{r=1}^R c_{mr}(Q_{mr}) \\
& \left. + \sum_{n=1}^N c_{mn}(Q_{nm}) + \sum_{r=1}^R v_{im} CF_{im} Q_{mr} + \sum_{r=1}^R u_{om} EF_{om} Q_{mr} \right) \\
\text{s.t.} \quad & \sum_{r=1}^R Q_{mr} \leq \varphi_{sm} \sum_{s=1}^S Q_{sm} + \varphi_{nm} \sum_{n=1}^N Q_{nm}.
\end{aligned} \tag{3}$$

The optimal conditions of all producers are equivalent to the following corresponding variational inequalities:

$$\begin{aligned}
& \sum_{m=1}^M \sum_{r=1}^R \left( \frac{\partial c_{mr}(Q_{mr})}{\partial Q_{mr}} + v_{im} CF_{im} + u_{om} EF_{om} + \zeta_{1m} - \rho_{mr} \right) \times (Q_{mr} - Q_{mr}^*) + (Q_{sm} - Q_{sm}^*) \\
& \times \sum_{m=1}^M \sum_{s=1}^S \left( \frac{\partial f_{ms}(Q_{sm}, \varphi_{sm})}{\partial Q_{sm}} + \frac{\partial c_{ms}(Q_{sm})}{\partial Q_{sm}} + \rho_{sm} - \varphi_{sm} \zeta_{1m} \right) + \sum_{m=1}^M \sum_{n=1}^N \left( \frac{\partial f_{mn}(Q_{nm}, \varphi_{nm})}{\partial Q_{nm}} + \frac{\partial c_{mn}(Q_{nm})}{\partial Q_{nm}} + \rho_{nm} - \varphi_{nm} \zeta_{1m} \right) \\
& \times (Q_{nm} - Q_{nm}^*) + \sum_{m=1}^M \left( \varphi_{sm} \sum_{s=1}^S Q_{sm} + \varphi_{nm} \sum_{n=1}^N Q_{nm} - \sum_{r=1}^R Q_{mr} \right) \times (\zeta_{1m} - \zeta_{1m}^*) \geq 0, \\
& \forall Q_{mr} \geq 0, Q_{sm} \geq 0, Q_{nm} \geq 0, \zeta_{1m} \geq 0,
\end{aligned} \tag{4}$$

where  $\zeta_{1m}$  in is the Lagrange coefficient to ensure the constraint condition of the objective function is established, which represents the minimum supply cost of the producer, that is, the cost that the producer is willing to bear at most for the unit product. The economic interpretation of variational inequality (4) is as follows.

The first one indicates that in equilibrium, for any trading volume greater than zero, the sum of the marginal transaction cost of the manufacturer and the retailer, the marginal energy use cost of the production input, the marginal carbon emission cost of the production output, and its minimum supply cost is equal to the transaction price of the auto parts manufacturer.

The second item indicates that in the equilibrium state, for any trading volume greater than zero, the marginal production cost of new materials, the marginal transaction cost of suppliers and producers, and the sum of the transaction price of suppliers is equal to the product of the availability coefficient and the minimum supply cost of suppliers' raw materials, and the third item is similar.

The fourth item indicates that in the equilibrium state, for any trading volume greater than zero, the sum of the trading volume of the auto parts manufacturer and the retailer is equal to the product of the trading volume of the supplier and the manufacturer and their availability coefficient of raw materials plus the product of the trading

volume of the recycler and the manufacturer and their availability coefficient of reprocessed products.

**3.1.3. Network Equilibrium Model of Physical Retailer and Online Retailer.** Suppose  $Q_{rd}$  is the physical retailer's and the online retailer's sales volume,  $\rho_{rd}$  is the physical retailer's and the online retailer's sales price, the unit energy

consumption cost is  $v_i$ , and the unit carbon emission cost is  $u_o$ . When the physical retailer's and the online retailer's cost is storage cost, payment to the manufacturer cost, and the transaction cost with the manufacturer and the demand market, plus energy consumption and the carbon emission cost in sales operation, then the objective function of the physical retailer and the online retailer  $r$  is

$$\begin{aligned} \max \quad & \sum_{d=1}^D \rho_{rd} Q_{rd} - \left( \sum_{m=1}^M c_r(Q_{mr}) + \sum_{m=1}^M \rho_{mr} Q_{mr} + \sum_{m=1}^M c_{rm}(Q_{mr}) + \sum_{d=1}^D c_{rd}(Q_{rd}) + \sum_{d=1}^D v_{ir} CF_{ir} Q_{rd} + \sum_{d=1}^D u_{or} EF_{or} Q_{rd} \right) \\ \text{s.t.} \quad & \sum_{d=1}^D Q_{rd} \leq \sum_{m=1}^M Q_{mr}. \end{aligned} \quad (5)$$

The optimal conditions of all physical retailers and online retailers are equivalent to the following variational inequalities:

$$\begin{aligned} & \sum_{r=1}^R \sum_{d=1}^D \left( \frac{\partial c_{rd}(Q_{rd})}{\partial Q_{rd}} + v_{ir} CF_{ir} + u_{or} EF_{or} + \zeta_{2r} - \rho_{rd} \right) \times (Q_{rd} - Q_{rd}^*) + \sum_{r=1}^R \sum_{m=1}^M \left( \frac{\partial c_r(Q_{mr})}{\partial Q_{mr}} + \frac{\partial c_{rm}(Q_{mr})}{\partial Q_{mr}} + \rho_{mr} - \zeta_{2r} \right) \\ & \times (Q_{mr} - Q_{mr}^*) + \sum_{r=1}^R \left( \sum_{m=1}^M Q_{mr} - \sum_{d=1}^D Q_{rd} \right) \times (\zeta_{2r} - \zeta_{2r}^*) \geq 0, \end{aligned}$$

$$\forall Q_{rd} \geq 0, Q_{mr} \geq 0, \zeta_{2r} \geq 0,$$

(6)

where the “\*” sign indicates that the variable is in equilibrium, corresponding to the optimal solution of variational inequality (6). Among them, the Lagrange coefficient  $\zeta_{2r}$  ensures that the constraint condition of the objective function holds, which means the physical retailer's and the online retailer's minimum supply cost, that is, the highest cost that the physical retailer and the online retailer are willing to pay when selling each unit of product. The variational inequality (6) shows the following.

The first one indicates that in equilibrium, for any trading volume greater than zero, the sum of the marginal transaction cost, the input energy use cost, the output unit carbon emission cost, and the minimum supply cost of the retailer is equal to the transaction price of the auto parts retailer.

The second indicates that in equilibrium, for any supply greater than zero, the sum of the retailer's marginal storage cost, the retailer's marginal transaction cost, and the manufacturer's transaction price is equal to the minimum supply cost of the auto parts retailer.

The third indicates that in equilibrium, for any supply greater than zero, the sum of the transaction volume of auto parts manufacturers and physical retailers and online retailers is equal to that of retailers and the demand market.

**3.1.4. Network Equilibrium Model of Demand Market.** The variables that the demand market needs to decide include two parts. In positive logistics, the demand market meets the following conditions:

$$\rho_{rd} + c_{dr}(Q_{rd}) \begin{cases} = \rho_d, & Q_{rd} > 0, \\ \geq \rho_d, & Q_{rd} = 0. \end{cases} \quad (7)$$

The demand market has the following supply-demand balance relationship:

$$D_d(\rho_d) \begin{cases} = \sum_{r=1}^R Q_{rd}, & \rho_d > 0, \\ \geq \sum_{r=1}^R Q_{rd}, & \rho_d = 0. \end{cases} \quad (8)$$

The economic explanation is as follows: formula (7) indicates that if the sum of the physical retailer's and the online retailer's selling price and the transaction cost of the demand market is equal to the demand price, then there is a transaction between the physical retailer, the online retailer, and the demand market; otherwise, the transaction volume between the two is zero; formula (8) indicates that if the demand of the demand market is exactly equal to the quantity of products purchased from the physical retailer and the online retailer, then the equilibrium price is positive; if the two are not equal, then the equilibrium demand price is zero.

In reverse logistics, the demand market  $D$  determines whether to sell to the recycler according to the recycling price of waste products. This relationship can be expressed as follows:

$$c_{dn}(Q_{dn}) \begin{cases} = \rho_{dn}, & Q_{dn} > 0, \\ \geq \rho_{dn}, & Q_{dn} = 0, \end{cases} \quad (9)$$

$$\text{s.t.} \quad \sum_{n=1}^N Q_{dn} \leq \sum_{r=1}^R Q_{rd}. \quad (10)$$

Equation (10) indicates that the total amount recovered by the recycler from the demand market  $D$  does not exceed the total amount of products in the demand market. By synthesizing the behavior and constraints of the demand market in forward logistics and reverse logistics, the optimal conditions of all demand markets can be obtained, which are equivalent to the following variational inequalities, and the solution  $(Q_{dn}, Q_{rd}, \rho_d, \zeta_{3d}) \geq 0$  satisfies

$$\begin{aligned} & \sum_{d=1}^D \sum_{r=1}^R [\rho_{rd} + c_{dr}(Q_{rd} - \rho_d - \zeta_{3d})] \times (Q_{rd} - Q_{rd}^*) + \sum_{d=1}^D \sum_{n=1}^N (c_{dn}(Q_{dn}) - \rho_{dn} + \zeta_{3d}) \times (Q_{dn} - Q_{dn}^*) \\ & + \sum_{d=1}^D \left( \sum_{r=1}^R Q_{rd} - D_d(\rho_d) \right) \times (\rho_d - \rho_d^*) + \sum_{d=1}^D \left( \sum_{r=1}^R Q_{rd} - \sum_{n=1}^N Q_{dn} \right) \times (\zeta_{3d} - \zeta_{3d}^*) \geq 0. \end{aligned} \quad (11)$$

**3.1.5. Network Equilibrium Model of Recycler.** It is assumed that the quantity of waste products sold by the recycler to the manufacturer is  $Q_{nm}$  and the selling price is  $\rho_{nm}$ .  $Q_{dn}$  is the total quantity of recycled waste products. The profit of the recycler is the profit obtained from the sale of reusable products to the manufacturer minus the cost of acquisition, transportation, and storage of the recycler, the cost of

recycling paid to the demand market, and the transaction cost with the demand market and the manufacturer, plus a part of the recycled waste products that need to be scrapped. The total amount of scrapping is  $(1 - \theta) \sum_{d=1}^D Q_{dn}$ , the cost is  $\chi(1 - \theta) \sum_{d=1}^D Q_{dn}$ , and the energy consumption and carbon emission in the recycling process cost, the objective function of recovery quotient  $n$ , is

$$\begin{aligned} \max \quad & \sum_{m=1}^M \rho_{nm} Q_{nm} - \left( \sum_{d=1}^D c_n(Q_{dn}) + \sum_{d=1}^D \rho_{dn} Q_{dn} + \sum_{m=1}^M c_{nm}(Q_{nm}) + \sum_{d=1}^D c_{nd}(Q_{nd}) + \chi(1 - \theta) \sum_{d=1}^D Q_{dn} + \sum_{d=1}^D v_{in} CF_{in} Q_{dn} + \sum_{d=1}^D u_{on} EF_{on} Q_{dn} \right) \\ \text{s.t.} \quad & \sum_{m=1}^M Q_{nm} \leq \theta \sum_{d=1}^D Q_{dn}. \end{aligned} \quad (12)$$

The optimal conditions of all recovery quotients are equivalent to the following variational inequalities:

$$\begin{aligned} & \sum_{n=1}^N \sum_{m=1}^M \left( \frac{\partial c_{nm}(Q_{nm})}{\partial Q_{nm}} + \zeta_{4n} - \rho_{nm} \right) \times (Q_{nm} - Q_{nm}^*) + \sum_{n=1}^N \sum_{d=1}^D \left( \frac{\partial c_n(Q_{dn})}{\partial Q_{dn}} + \frac{\partial c_{nd}(Q_{dn})}{\partial Q_{dn}} + \rho_{dn} + v_{in} CF_{in} + u_{on} EF_{on} + \chi((1 - \theta) - \theta \zeta_{4n}) \right) \\ & \times (Q_{dn} - Q_{dn}^*) + \sum_{n=1}^N \left( \theta \sum_{d=1}^D Q_{dn} - \sum_{m=1}^M Q_{nm} \right) \times (\zeta_{4n} - \zeta_{4n}^*) \geq 0, \forall Q_{nm} \geq 0, Q_{dn} \geq 0, \zeta_{4n} \geq 0, \end{aligned} \quad (13)$$

where  $\zeta_{4n}$  is the Lagrangian coefficient to ensure the constraint condition of the objective function is established, which represents the minimum recovery cost of the recycler, that is, the cost that the recycler is willing to bear at most for the unit product recovery. The economic explanation of variational inequality (13) is as follows.

The first item indicates that in equilibrium, for any supply greater than zero, the marginal transaction price of the auto parts recycler is equal to the sum of the marginal transaction cost of the recycler and the manufacturer and the minimum recovery cost of the recycler.

The second item indicates that in the equilibrium state, for any supply greater than zero, the sum of the marginal storage cost of the recycler, the marginal transaction cost of the recycler and the demand market, the transaction price of the demand market, the energy use cost of the input end, the carbon emission cost of the output end, and the disposal cost of the unit scrap product is equal to the minimum recovery cost of the auto parts recycler.

The third item indicates that under the equilibrium state, for any supply greater than zero, the transaction volume of

the auto parts demand market and the recycler is equal to that of the recycler and the manufacturer.

**3.1.6. Network Equilibrium Model with Carbon Trading as Objective Function.** In the supply chain, when the carbon emission limit of  $A_{CC}$  node enterprise  $j$  is not enough, it can buy the carbon emission limit  $CC_j^-$  sold by other enterprises at a price through the carbon trading market; on the contrary, when the carbon emission limit of the enterprise with good energy conservation and emission reduction is redundant, it can sell the carbon emission limit of the enterprise  $CC_j^+$  needed at  $V^{CC}$  price. Carbon emissions are calculated in terms of carbon dioxide equivalent per ton ( $CO_2e$ ). Then, the objective function of the trading profit of a node enterprise  $j$  is

$$\text{Max } V^{CC}CC_j^+ - A^{CC}CC_j^-, \quad (14)$$

where  $j \in M \cup R \cup N$ .

The constraint is that the energy consumption and the carbon footprint of each node enterprise in the supply chain network are within the limit, that is,

$$\begin{aligned} CO_o^{\text{out}} \sum_{r=1}^R EF_{om} Q_{mr} + CO_i^{\text{in}} \sum_{r=1}^R CF_{im} Q_{mr} + CC_m^+ - CC_m^- &\leq L_m^{\text{CO}_2e}, \\ CO_o^{\text{out}} \sum_{d=1}^D EF_{or} Q_{rd} + CO_i^{\text{in}} \sum_{d=1}^D CF_{ir} Q_{rd} + CC_r^+ - CC_r^- &\leq L_r^{\text{CO}_2e}, \\ CO_o^{\text{out}} \sum_{d=1}^D EF_{on} Q_{dn} + CO_i^{\text{in}} \sum_{d=1}^D CF_{in} Q_{dn} + CC_n^+ - CC_n^- &\leq L_n^{\text{CO}_2e}, \\ CC_m^+ &\leq L_m^{\text{COS}}, \\ CC_m^- &\leq L_m^{\text{COP}}, \\ CC_r^+ &\leq L_r^{\text{COS}}, \\ CC_r^- &\leq L_r^{\text{COP}}, \\ CC_n^+ &\leq L_n^{\text{COS}}, \\ CC_n^- &\leq L_n^{\text{COP}}. \end{aligned} \quad (15)$$

The solution of the objective function with constraints on the carbon footprint of the whole low-carbon supply chain is equivalent to the solution of the corresponding

variational inequality below, which satisfies the following conditions:

$$\begin{aligned}
& (CC_m^+, CC_m^-, CC_r^+, CC_r^-, CC_n^+, CC_n^-, Q_{mr}, Q_{rd}, Q_{dn}, \zeta_{5m}, \xi_{5r}, \tau_{5n}, \alpha_{5m}, \alpha'_{5m}, \beta_{5r}, \beta'_{5r}, \gamma_{5n}, \gamma'_{5n}) \geq 0 \\
& \sum_{m=1}^M (\zeta_{5m} + \alpha_{5m} - V^{cc}) \times (CC_m^+ - CC_m^{+*}) + \sum_{m=1}^M (A^{cc} - \zeta_{5m} + \alpha'_{5m}) \times (CC_m^- - CC_m^{-*}) \\
& + \sum_{r=1}^R (\xi_{5r} + \beta_{5r} - V^{cc}) \times (CC_r^+ - CC_r^{+*}) + \sum_{r=1}^R (A^{cc} - \xi_{5r} + \beta'_{5r}) \times (CC_r^- - CC_r^{-*}) \\
& + \sum_{n=1}^N (\tau_{5n} + \gamma_{5n} - V^{cc}) \times (CC_n^+ - CC_n^{+*}) + \sum_{n=1}^N (A^{cc} - \tau_{5n} + \gamma'_{5n}) \times (CC_n^- - CC_n^{-*}) \\
& + \sum_{m=1}^M \sum_{r=1}^R (CO_o^{\text{out}} EF_{om} + CO_i^{\text{in}} CF_{im}) \zeta_{5m} \times (Q_{mr} - Q_{mr}^*) \\
& + \sum_{r=1}^R \sum_{d=1}^D (CO_o^{\text{out}} EF_{or} + CO_i^{\text{in}} CF_{ir}) \xi_{5r} \times (Q_{rd} - Q_{rd}^*) \\
& + \sum_{n=1}^N \sum_{d=1}^D (CO_o^{\text{out}} EF_{on} + CO_i^{\text{in}} CF_{in}) \tau_{5n} \times (Q_{dn} - Q_{dn}^*) \\
& + \sum_{m=1}^M \left[ L_m^{\text{CO}_2^e} - CO_o^{\text{out}} \sum_{r=1}^R EF_{om} Q_{mr} - CO_i^{\text{in}} \sum_{r=1}^R CF_{im} Q_{mr} - CC_m^+ + CC_m^- \right] \times (\zeta_{5m} - \zeta_{5m}^*) \\
& + \sum_{r=1}^R \left[ L_r^{\text{CO}_2^e} - CO_o^{\text{out}} \sum_{d=1}^D EF_{or} Q_{rd} - CO_i^{\text{in}} \sum_{d=1}^D CF_{ir} Q_{rd} - CC_r^+ + CC_r^- \right] \times (\xi_{5r} - \xi_{5r}^*) \\
& + \sum_{n=1}^N \left[ L_n^{\text{CO}_2^e} - CO_o^{\text{out}} \sum_{d=1}^D EF_{on} Q_{dn} - CO_i^{\text{in}} \sum_{d=1}^D CF_{in} Q_{dn} - CC_n^+ + CC_n^- \right] \times (\tau_{5n} - \tau_{5n}^*) \\
& + \sum_{m=1}^M (L_m^{\text{COS}} - CC_m^+) \times (\alpha_{5m} - \alpha_{5m}^*) + \sum_{m=1}^M (L_m^{\text{COP}} - CC_m^-) \times (\alpha'_{5m} - \alpha'_{5m}^*) \\
& + \sum_{r=1}^R (L_r^{\text{COS}} - CC_r^+) \times (\beta_{5r} - \beta_{5r}^*) + \sum_{r=1}^R (L_r^{\text{COP}} - CC_r^-) \times (\beta'_{5r} - \beta'_{5r}^*) \\
& + \sum_{n=1}^N (L_n^{\text{COS}} - CC_n^+) \times (\gamma_{5n} - \gamma_{5n}^*) + \sum_{n=1}^N (L_n^{\text{COP}} - CC_n^-) \times (\gamma'_{5n} - \gamma'_{5n}^*) \geq 0,
\end{aligned} \tag{16}$$

where  $\zeta_{5m}, \xi_{5r}, \tau_{5n}, \alpha_{5m}, \alpha'_{5m}, \beta_{5r}, \beta'_{5r}, \gamma_{5n}, \gamma'_{5n}$  is the Lagrange coefficient to ensure the constraint conditions of the objective function is established.

**3.2. The Establishment of Network Equilibrium Model under the Environment Goal of Low-Carbon Supply Chain.** Based on the carbon footprint management, the environmental objective function of the low-carbon supply chain is minimum carbon emission. Taking the carbon dioxide equivalent per ton tCO<sub>2</sub>e as the calculation unit, the environmental objective function of the whole low-carbon supply chain network is

$$\min CO_i^{\text{in}} C_i + CO_o^{\text{out}} E_o. \tag{17}$$

The total input energy consumption and output carbon emissions are as follows:

$$\begin{aligned}
C_i &= \sum_{m=1}^M \sum_{r=1}^R CF_{im} Q_{mr} + \sum_{r=1}^R \sum_{d=1}^D CF_{ir} Q_{rd} + \sum_{n=1}^N \sum_{d=1}^D CF_{in} Q_{dn}, \\
E_o &= \sum_{m=1}^M \sum_{r=1}^R EF_{om} Q_{mr} + \sum_{r=1}^R \sum_{d=1}^D EF_{or} Q_{rd} + \sum_{n=1}^N \sum_{d=1}^D EF_{on} Q_{dn}.
\end{aligned} \tag{18}$$

The optimal conditions of the above programming problems are equivalent to the following variational inequalities, and the solution  $(Q_{mr}, Q_{rd}, Q_{dn}) \geq 0$  satisfies the following requirements:



$$\begin{aligned}
& \sum_{m=1}^M \sum_{r=1}^R (CO_o^{\text{out}} EF_{om} + CO_i^{\text{in}} CF_{im}) \times (Q_{mr} - Q_{mr}^*) \\
& + \sum_{r=1}^R \sum_{d=1}^D (CO_o^{\text{out}} EF_{or} + CO_i^{\text{in}} CF_{ir}) \times (Q_{rd} - Q_{rd}^*) \\
& + \sum_{n=1}^N \sum_{d=1}^D (CO_o^{\text{out}} EF_{on} + CO_i^{\text{in}} CF_{in}) \times (Q_{dn} - Q_{dn}^*) \geq 0.
\end{aligned} \tag{19}$$

**3.3. The Network Equilibrium Model of the Whole Low-Carbon Supply Chain.** The quantity of purchase and recycled products in the demand market must be equal to the quantity received by the reprocessing recycler. The price and the carbon trading volume between the upstream and downstream networks of the supply chain must meet the sum of the equivalent variational inequalities of each equilibrium condition. When the low-carbon supply chain network finally reaches the equilibrium state, its decision variables must meet the optimal conditions of suppliers, producers, physical retailers, online retailers, demand markets, and recyclers, as well as the optimal conditions of carbon trading and carbon footprint environmental objective function.

## 4. Simulation and Result Analysis

In this section, the above models are simulated to discuss the change rule of trading profit and carbon emission with exogenous variables (recycling coefficient, remanufacturing coefficient of waste products, etc.). It is assumed that the supply chain network model of auto parts is a closed-loop supply chain composed of two suppliers, two producers, one physical retailer, online retailer, two demand markets, and two recyclers. For the purpose of simulation calculation, the parameters used in the model and the settings of the functions involved in the model are shown in Table 3. There are many parameters involved in the model. It mainly analyzes the recycling coefficient  $r_k$  and the remanufacturing coefficient  $\mu$  of the network members and observes the changes of the total profits of carbon trading and the total carbon emissions of the supply chain.

**4.1. The Influence of Recycling Coefficient on Carbon Emission and Total Profit of Supply Chain.** Under the different recycling coefficient  $r_k$ , the total carbon emission and the total profit of carbon trading and selling in the supply chain show an interval change trend, as shown in Figures 2 and 3, respectively. In order to facilitate the analysis of the impact of the change of the recycling coefficient on the carbon emissions and total profits of the supply chain, this paper considers the following.

It can be seen from Figure 2 that with the increase in the recycle coefficient  $r_k$ , the total amount of carbon emissions increases continuously and the range is large. This is because the larger the  $r_k$  is in the process of recycling waste products,

TABLE 3: Parameter setting.

| Parameter           | Value             |
|---------------------|-------------------|
| $v_{im}$            | 200/t             |
| $v_{ir}$            | 200/t             |
| $v_{in}$            | 200/t             |
| $A^{cc}$            | 2500/t            |
| $V^{cc}$            | 2500/t            |
| $a$                 | 0.5               |
| $g$                 | 4                 |
| $p$                 | 1000              |
| $u_{om}$            | 200/t             |
| $u_{or}$            | 200/t             |
| $u_{on}$            | 200/t             |
| $\varphi_{sm}$      | 0.9               |
| $\varphi_{nm}$      | 0.8               |
| $b$                 | 2                 |
| $h$                 | 1                 |
| $q$                 | 0.4               |
| $CF_{im}$           | 6                 |
| $CF_{ir}$           | 3                 |
| $CF_{in}$           | 8                 |
| $CO_i^{\text{in}}$  | 0.015             |
| $CO_o^{\text{out}}$ | 0.015             |
| $c$                 | 0.8               |
| $k$                 | 3                 |
| $t$                 | 0.02              |
| $EF_{om}$           | 12                |
| $EF_{or}$           | 6                 |
| $EF_{on}$           | 15                |
| $\chi$              | 2                 |
| $\lambda$           | $10^{-2}$         |
| $e$                 | 3.5               |
| $l$                 | 2.5               |
| $u$                 | 0.03              |
| $L^{\text{COS}}$    | $8 \times 10^5/t$ |
| $L^{\text{COP}}$    | $8 \times 10^5/t$ |
| $L^{\text{CO}_2e}$  | $8 \times 10^5/t$ |
| $\varepsilon$       | $10^{-3}$         |
| $\theta$            | 0.8               |
| $f$                 | 5                 |
| $o$                 | 1.5               |

the more the waste products will be recycled from the demand market, the more the waste products will be disposed, and the more the carbon emissions will be.

It can be seen from Figure 3 that with the increase in the recycling coefficient  $r_k$ , recycling conversion and reuse coefficient  $\theta$  shows an upward trend, while the trading profit of carbon trading shows a downward trend. It shows that the increase of recycled products recycling leads to the increase of recycled products conversion and reuse, which increases the carbon emissions of supply chain members, while the carbon emission quota allocated by most members of the supply chain is not enough to make up for the increased carbon emissions and reduced carbon trading, thus reducing the trading profits of carbon trading.

**4.2. Influence of Remanufacturing Coefficient of Waste Products on Carbon Emission and Total Profit of Supply Chain.** With different remanufacturing coefficient  $\mu$  of waste products, the total amount of carbon emissions in the

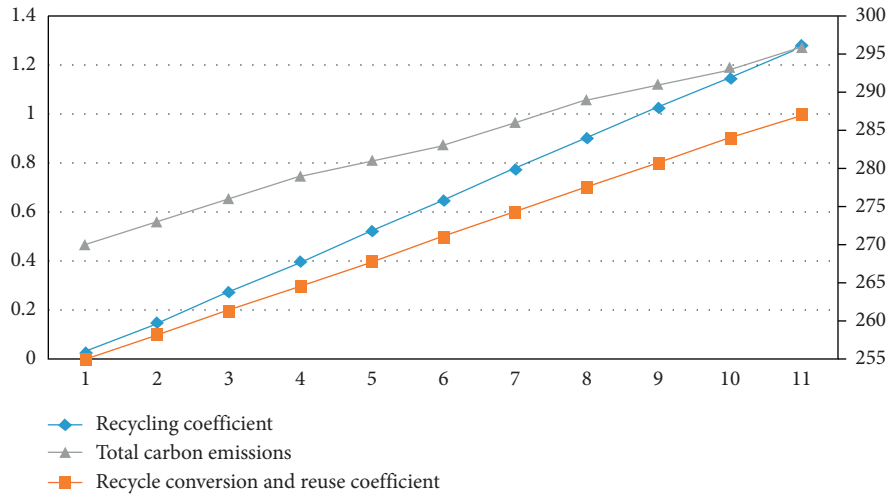


FIGURE 2: Change of total carbon emission with  $r_k$ .

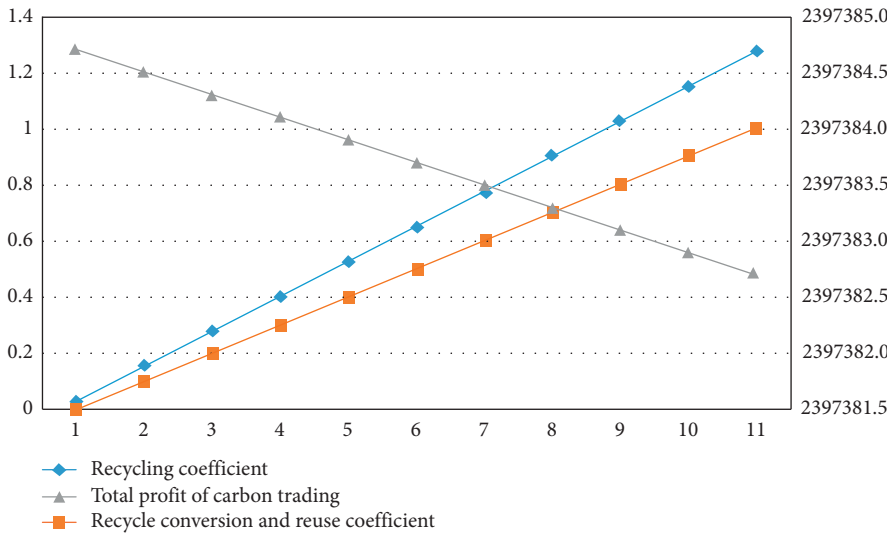


FIGURE 3: Change of total profit of carbon trading with  $r_k$ .

supply chain and the total profit of carbon trading show an interval trend, as shown in Figures 4 and 5, respectively. In order to analyze the influence of the remanufacturing coefficient of waste products on carbon emission and total profit of the supply chain, this paper considers the following.

As can be seen from Figure 4, the carbon emissions in the supply chain increase with the increase in  $\mu$ , but the range is not very large. This is due to the increase of recycled materials and the corresponding increase of production, thus the carbon emissions increase. Recyclers have a great impact on carbon emissions. By properly increasing the remanufacturing coefficient of waste products, carbon emissions and trading profits can be optimized at the same time.

As can be seen from Figure 5, with the increase in the remanufacturing coefficient  $\mu$  of waste products, the availability coefficient of reprocessed products of recyclers is also slowly increasing and the trading profit of carbon trading decreases accordingly. This is because with the manufacturing of waste products, recyclers can process more products, and the carbon emission quota allocated by producers is not enough to make up for the increased carbon emissions, which reduces the carbon trading and thus reduces the profits of carbon trading. In order to improve the profit of carbon trading in the supply chain, manufacturers need to make a reasonable ordering plan in order to balance the carbon emissions and the profit of carbon trading.

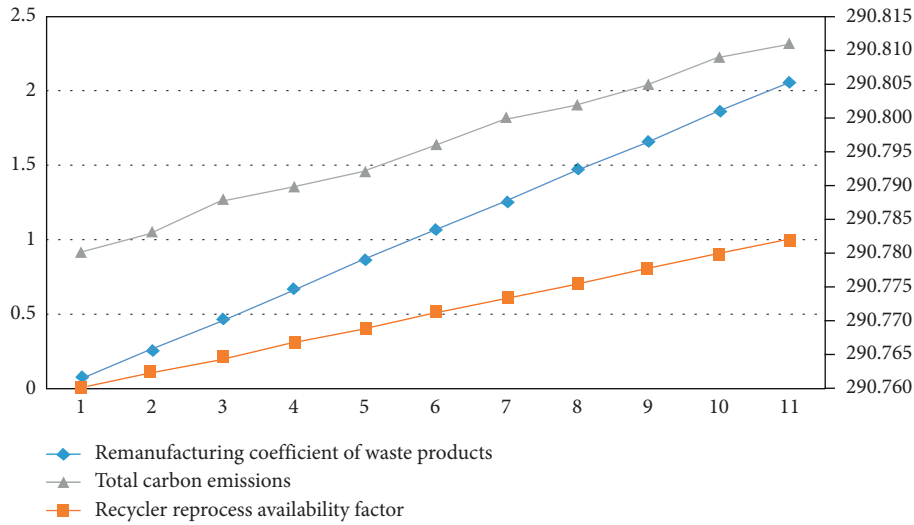


FIGURE 4: Change of total carbon emission with the remanufacturing coefficient  $\mu$  of waste products.

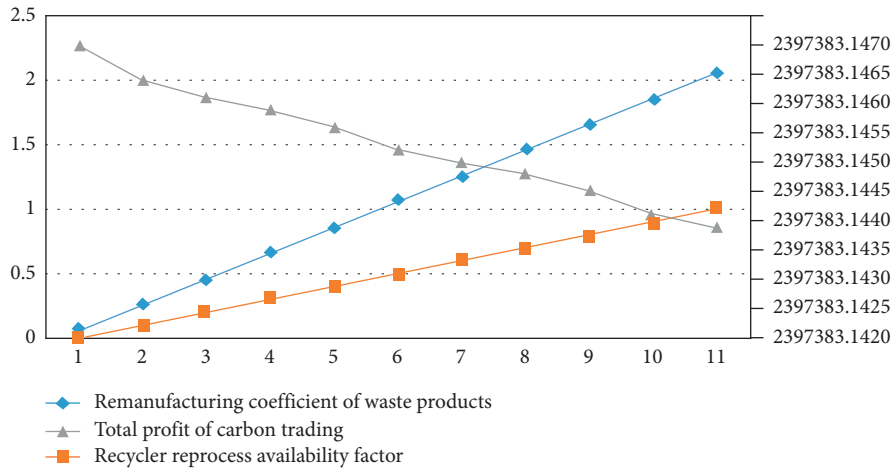


FIGURE 5: Change of total profit of carbon trading with the remanufacturing coefficient  $\mu$  of waste products.

### 5. Conclusion

The top-level design of a smart city should focus on green in strategy, aiming to build a green smart city. Smart cities plan and build around energy conservation, emission reduction, and environmental optimization and regard sustainable development as the starting point and destination. As an important strategic industry in the construction of smart cities, the auto parts industry should also take responsibility for low-carbon emission reduction. In this paper, by considering the optimal decision-making behavior of suppliers, producers, physical retailers, online retailers, demand markets, and recyclers, as well as the situation of carbon trading and carbon emissions, the equilibrium conditions of each node in the low-carbon supply chain of auto parts are analyzed by using variational inequality, and the network equilibrium optimization model of the low-carbon supply chain is established. Using the simulation analysis method, through analyzing the relationship between carbon emissions and trading profits with the recovery coefficient of

recycled products and the remanufacturing coefficient of waste products, we can get that carbon emissions directly affect the trading profits of carbon trading; recyclers have a great impact on carbon emissions. Properly increasing the recovery coefficient of recycled products and the remanufacturing coefficient of waste products can make carbon emissions and profits of carbon trading reach the optimal level at the same time. In order to improve the profits of carbon trading in the supply chain, manufacturers need to make reasonable ordering plans to balance carbon emissions and profits of carbon trading. It further shows that in order to achieve the network equilibrium of the low-carbon supply chain of auto parts, the coordination and cooperation of supply chain members is needed. Through recycling, process improvement, and other measures, the carbon emissions of the supply chain network can be reduced, and the excess carbon emissions can be sold in the carbon trading market. The profits can offset the cost investment in emission reduction and finally achieve the purpose of low-carbon supply chain management.

## Data Availability

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also form a part of an ongoing study.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The funding was sponsored by the National Social Science Fund of China (Grant no. 18CGL015).

## References

- [1] J. Huang, S. B. Feng, and Y. T. Niu, "Influence of smart city on green low carbon development," *On Economic Problems*, vol. 477, no. 5, pp. 128–135, 2019.
- [2] S. Y. Ren, H. C. Dai, P. Wang, D. Q. Zhao, and T. Masui, "Economic impacts of carbon emission trading: case study on Guangdong province," *Advances in Climate Change Research*, vol. 11, no. 1, pp. 61–67, 2015.
- [3] S. X. Ma, Y. Wang, and M. Z. Yang, "Urban construction of low-carbon eco-city from the perspective of smart city," *Journal of Shenyang Jianzhu University (Social Science)*, vol. 20, no. 6, pp. 40–44, 2018.
- [4] J. Yin and J. Liu, "Research on low-carbon paths of automotive supply chain," *Ecological Economy*, vol. 30, no. 11, pp. 57–60, 2014.
- [5] K.-H. Lee, "Integrating carbon footprint into supply chain management: the case of Hyundai Motor Company (HMC) in the automobile industry," *Journal of Cleaner Production*, vol. 19, no. 11, pp. 1216–1223, 2011.
- [6] L. Yang and Z. Guitao, "The closed-loop supply chain network decision based on corporate environmental criteria and consumer's environmental consciousness," *China Population, Resource and Environment*, vol. 29, no. 11, pp. 71–81, 2019.
- [7] S. D. Wei, "Differential game analysis on supply chain's strategy of carbon emission reduction R & D under carbon trading policy," *Chinese Journal of Management Science*, vol. 15, no. 5, pp. 782–790, 2018.
- [8] M. M. Wang and X. L. Han, "Location-routing-inventory joint optimization of perishable product supply chain considering carbon emission," *Journal of Shanghai Maritime University*, vol. 40, no. 4, pp. 45–51, 2019.
- [9] L. A. Diaz-Trujillo, J. Tovar-Facio, F. Nápoles-Rivera, and J. M. Ponce-Ortega, "Effective use of carbon pricing on climate change mitigation projects: analysis of the biogas supply chain to substitute liquefied-petroleum gas in Mexico," *Processes*, vol. 7, no. 10, p. 668, 2019.
- [10] M. Efthymiou and A. Papatheodorou, "EU Emissions Trading scheme in aviation: policy analysis and suggestions," *Journal of Cleaner Production*, vol. 237, 2019.
- [11] D. Battini, M. Calzavara, I. Isolan, F. Sgarbossa, and F. Zangaro, "Sustainability in material purchasing: a multi-objective economic order quantity model under Carbon Trading," *Sustainability*, vol. 10, no. 12, 2018.
- [12] M. W. Zhang, X. L. Qu, and B. Li, "Coordination optimization in multi-product and multi-objective supply chains considering carbon emission," *Computer Integrated Manufacturing Systems*, vol. 24, no. 4, pp. 1024–1033, 2018.
- [13] Z. R. Lan, Z. Q. Sun, and F. Y. Jing, "Research on the dynamic production decision of enterprises under the constraint of carbon emission quota," *Statistics & Decision*, vol. 34, no. 24, pp. 66–69, 2018.
- [14] A. A. Taleizadeh, N. Alizadeh-Basban, and S. T. A. Niaki, "A closed-loop supply chain considering carbon reduction, quality improvement effort, and return policy under two remanufacturing scenarios," *Journal of Cleaner Production*, vol. 232, pp. 1230–1250, 2019.
- [15] I. Nouira, R. Hammami, Y. Frein, and C. Temponi, "Design of forward supply chains: impact of a carbon emissions-sensitive demand," *International Journal of Production Economics*, vol. 173, pp. 80–98, 2016.
- [16] Z. Huang, X. Xu, H. Zhu, and M. Zhou, "An efficient group recommendation model with multiattention-based neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, 2020.
- [17] A. K. Purohit, R. Shankar, P. K. Choudhary, and A. Choudhary, "Non-stationary stochastic inventory lot-sizing with emission and service level constraints in a carbon cap-and-trade system," *Journal of Cleaner Production*, vol. 113, pp. 654–661, 2016.
- [18] L. K. Saxena, P. K. Jain, and A. K. Sharma, "Tactical supply chain planning for tyre remanufacturing considering carbon tax policy," *International Journal of Advanced Manufacturing Technology*, vol. 97, no. 1–4, pp. 1505–1528, 2018.
- [19] C. Mauro, A. A. Rentizelas, and D. Chinese, "International vs. domestic bioenergy supply chains for co-firing plants: the role of pre-treatment technologies," *Renewable Energy*, vol. 119, pp. 712–730, 2018.
- [20] J. Di Filippo, J. Karpman, and J. R. DeShazo, "The impacts of policies to reduce CO<sub>2</sub> emissions within the concrete supply chain," *Cement and Concrete Composites*, vol. 101, pp. 67–82, 2019.
- [21] J. Hu, Y. Sun, G. Li, G. Jiang, and B. Tao, "Probability analysis for grasp planning facing the field of medical robotics," *Measurement*, vol. 141, pp. 227–234, 2019.
- [22] L. Roibás, S. Rodríguez-García, V. P. Valdramidis, and A. Hospido, "The relevance of supply chain characteristics in GHG emissions: the carbon footprint of Maltese juices," *Food Research International*, vol. 107, pp. 747–754, 2018.
- [23] I. A. G. Wilson and I. Staffell, "Rapid fuel switching from coal to natural gas through effective carbon pricing," *Nature Energy*, vol. 3, no. 5, pp. 365–372, 2018.
- [24] E. Allevi, A. Gnudi, I. V. Konnov, and G. Oggioni, "Evaluating the effects of environmental regulations on a closed-loop supply chain network: a variational inequality approach," *Annals of Operations Research*, vol. 261, no. 1–2, pp. 1–43, 2018.
- [25] Z. Huang, X. Xu, J. Ni, H. Zhu, and C. Wang, "Multimodal representation learning for recommendation in Internet of Things," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10675–10685, 2019.
- [26] K. Shaw, M. Irfan, R. Shankar, and S. S. Yadav, "Low carbon chance constrained supply chain network design problem: a Benders decomposition based approach," *Computers & Industrial Engineering*, vol. 98, pp. 483–497, 2016.
- [27] V. F. Yu, R. Maglasang, and Y.-C. Tsao, "Shelf space allocation problem under carbon tax and emission trading policies," *Journal of Cleaner Production*, vol. 196, pp. 438–451, 2018.
- [28] A. Singh, S. Kumari, H. Malekpoor, and N. Mishra, "Big data cloud computing framework for low carbon supplier selection in the beef supply chain," *Journal of Cleaner Production*, vol. 202, pp. 139–149, 2018.