

Intelligent and Holistic Internet of Things (IoT) Solutions for Sustainable Cities and Society

Lead Guest Editor: Bo Rong

Guest Editors: Shengjie Xu, Michel Kadoch, and Peng Yu





**Intelligent and Holistic Internet of Things
(IoT) Solutions for Sustainable Cities and
Society**

Wireless Communications and Mobile Computing

Intelligent and Holistic Internet of Things (IoT) Solutions for Sustainable Cities and Society

Lead Guest Editor: Bo Rong

Guest Editors: Shengjie Xu, Michel Kadoch, and Peng Yu




Copyright © 2021 Hindawi Limited. All rights reserved.

This is a special issue published in “Wireless Communications and Mobile Computing.” 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

































Zhipeng Cai , USA

Associate Editors

Ke Guan , China
Jaime Lloret , Spain
Maode Ma , Singapore

Academic Editors

Muhammad Inam Abbasi, Malaysia
Ghufran Ahmed , Pakistan
Hamza Mohammed Ridha Al-Khafaji ,
Iraq
Abdullah Alamoodi , Malaysia
Marica Amadeo, Italy
Sandhya Aneja, USA
Mohd Dilshad Ansari, India
Eva Antonino-Daviu , Spain
Mehmet Emin Aydin, United Kingdom
Parameshchhari B. D. , India
Kalapraveen Bagadi , India
Ashish Bagwari , India
Dr. Abdul Basit , Pakistan
Alessandro Bazzi , Italy
Zdenek Becvar , Czech Republic
Nabil Benamar , Morocco
Olivier Berder, France
Petros S. Bithas, Greece
Dario Bruneo , Italy
Jun Cai, Canada
Xuesong Cai, Denmark
Gerardo Canfora , Italy
Rolando Carrasco, United Kingdom
Vicente Casares-Giner , Spain
Brijesh Chaurasia, India
Lin Chen , France
Xianfu Chen , Finland
Hui Cheng , United Kingdom
Hsin-Hung Cho, Taiwan
Ernestina Cianca , Italy
Marta Cimitile , Italy
Riccardo Colella , Italy
Mario Collotta , Italy
Massimo Condoluci , Sweden
Antonino Crivello , Italy
Antonio De Domenico , France
Floriano De Rango , Italy


Antonio De la Oliva , Spain
Margot Deruyck, Belgium
Liang Dong , USA
Praveen Kumar Donta, Austria
Zhuojun Duan, USA
Mohammed El-Hajjar , United Kingdom
Oscar Esparza , Spain
Maria Fazio , Italy
Mauro Femminella , Italy
Manuel Fernandez-Veiga , Spain
Gianluigi Ferrari , Italy
Luca Foschini , Italy
Alexandros G. Fragkiadakis , Greece
Ivan Ganchev , Bulgaria
Óscar García, Spain
Manuel García Sánchez , Spain
L. J. García Villalba , Spain
Miguel Garcia-Pineda , Spain
Piedad Garrido , Spain
Michele Girolami, Italy
Mariusz Glabowski , Poland
Carles Gomez , Spain
Antonio Guerrieri , Italy
Barbara Guidi , Italy
Rami Hamdi, Qatar
Tao Han, USA
Sherief Hashima , Egypt
Mahmoud Hassaballah , Egypt
Yejun He , China
Yixin He, China
Andrej Hrovat , Slovenia
Chunqiang Hu , China
Xuexian Hu , China
Zhenghua Huang , China
Xiaohong Jiang , Japan
Vicente Julian , Spain
Rajesh Kaluri , India
Dimitrios Katsaros, Greece
Muhammad Asghar Khan, Pakistan
Rahim Khan , Pakistan
Ahmed Khattab, Egypt
Hasan Ali Khattak, Pakistan
Mario Kolberg , United Kingdom
Meet Kumari, India
Wen-Cheng Lai , Taiwan

Jose M. Lanza-Gutierrez, Spain
Pavlos I. Lazaridis , United Kingdom
Kim-Hung Le , Vietnam
Tuan Anh Le , United Kingdom
Xianfu Lei, China
Jianfeng Li , China
Xiangxue Li , China
Yaguang Lin , China
Zhi Lin , China
Liu Liu , China
Mingqian Liu , China
Zhi Liu, Japan
Miguel López-Benítez , United Kingdom
Chuanwen Luo , China
Lu Lv, China
Basem M. ElHalawany , Egypt
Imadeldin Mahgoub , USA
Rajesh Manoharan , India
Davide Mattera , Italy
Michael McGuire , Canada
Weizhi Meng , Denmark
Klaus Moessner , United Kingdom
Simone Morosi , Italy
Amrit Mukherjee, Czech Republic
Shahid Mumtaz , Portugal
Giovanni Nardini , Italy
Tuan M. Nguyen , Vietnam
Petros Nicolitidis , Greece
Rajendran Parthiban , Malaysia
Giovanni Pau , Italy
Matteo Petracca , Italy
Marco Picone , Italy
Daniele Pinchera , Italy
Giuseppe Piro , Italy
Javier Prieto , Spain
Umair Rafique, Finland
Maheswar Rajagopal , India
Sujan Rajbhandari , United Kingdom
Rajib Rana, Australia
Luca Reggiani , Italy
Daniel G. Reina , Spain
Bo Rong , Canada
Mangal Sain , Republic of Korea
Praneet Saurabh , India

Hans Schotten, Germany
Patrick Seeling , USA
Muhammad Shafiq , China
Zaffar Ahmed Shaikh , Pakistan
Vishal Sharma , United Kingdom
Kaize Shi , Australia
Chakchai So-In, Thailand
Enrique Stevens-Navarro , Mexico
Sangeetha Subbaraj , India
Tien-Wen Sung, Taiwan
Suhua Tang , Japan
Pan Tang , China
Pierre-Martin Tardif , Canada
Sreenath Reddy Thummaluru, India
Tran Trung Duy , Vietnam
Fan-Hsun Tseng, Taiwan
S Velliangiri , India
Quoc-Tuan Vien , United Kingdom
Enrico M. Vitucci , Italy
Shaohua Wan , China
Dawei Wang, China
Huaqun Wang , China
Pengfei Wang , China
Dapeng Wu , China
Huaming Wu , China
Ding Xu , China
YAN YAO , China
Jie Yang, USA
Long Yang , China
Qiang Ye , Canada
Changyan Yi , China
Ya-Ju Yu , Taiwan
Marat V. Yuldashev , Finland
Sherali Zeadally, USA
Hong-Hai Zhang, USA
Jiliang Zhang, China
Lei Zhang, Spain
Wence Zhang , China
Yushu Zhang, China
Kechen Zheng, China
Fuhui Zhou , USA
Meiling Zhu, United Kingdom
Zhengyu Zhu , China

Contents

Industrial Internet Federated Learning Driven by IoT Equipment ID and Blockchain

Xu Zhang , Haibo Hou, Zhao Fang, and Zhiqian Wang


Research Article (9 pages), Article ID 7705843, Volume 2021 (2021)

Informatization Construction of Digital Assets in Smart Cities

Wenze Ning  and Mei Lu



Research Article (9 pages), Article ID 3344987, Volume 2021 (2021)

Evolution of Regional Economic Spatial Structure Based on IoT and GIS Service

Lei Jiang 

Research Article (11 pages), Article ID 6501865, Volume 2021 (2021)

BEI-TAB: Enabling Secure and Distributed Airport Baggage Tracking with Hybrid Blockchain-Edge System

Pengbo Si , Fei Wang , Enchang Sun, and Yuzhao Su


Research Article (12 pages), Article ID 2741435, Volume 2021 (2021)

A Weakly Supervised Academic Search Model Based on Knowledge-Enhanced Feature Representation

Mingying Xu , Junping Du , Feifei Kou , Meiyu Liang , Xin Xu , and Jiaxin Yang 

Research Article (12 pages), Article ID 4411524, Volume 2021 (2021)

Innovation and Development of Rural Leisure Tourism Industry Using Mobile Cloud IoT Computing

Guangwei Wang 





Research Article (11 pages), Article ID 6211063, Volume 2021 (2021)

Multitarget Real-Time Tracking Algorithm for UAV IoT

Tao Hong , Qiye Yang , Peng Wang, Jinmeng Zhang, Wenbo Sun, Lei Tao, Chaoqun Fang, and Jihan Cao



Research Article (15 pages), Article ID 9999596, Volume 2021 (2021)

Cooperative Offloading in D2D-Enabled Three-Tier MEC Networks for IoT

Jingyan Wu , Jiawei Zhang , Yuming Xiao , and Yuefeng Ji 


Research Article (13 pages), Article ID 9977700, Volume 2021 (2021)

CNN and DCGAN for Spectrum Sensors over Rayleigh Fading Channel

Junsheng Mu , Youheng Tan, Dongliang Xie, Fangpei Zhang, and Xiaojun Jing 

Research Article (12 pages), Article ID 9970600, Volume 2021 (2021)

Influencing Factors and Forecasting Statistics of Enterprise Market Sales Based on Big Data and Intelligent IoT


Zhen Guo and Tao Zou 

Research Article (13 pages), Article ID 9981581, Volume 2021 (2021)

Indoor PDR Positioning Assisted by Acoustic Source Localization, and Pedestrian Movement Behavior Recognition, Using a Dual-Microphone Smartphone

Mei Wang, Nan Duan, Zou Zhou , Fei Zheng , Hongbing Qiu , Xiaopeng Li, and Guoli Zhang
Research Article (16 pages), Article ID 9981802, Volume 2021 (2021)

Intelligent IoT-Based Cross-Border e-Commerce Supply Chain Performance Optimization

Lei Xia and Sitong Liu 
Research Article (13 pages), Article ID 9961925, Volume 2021 (2021)

6G Green IoT Network: Joint Design of Intelligent Reflective Surface and Ambient Backscatter Communication

Qiang Liu , Songlin Sun , Heng Wang , and Shaowei Zhang 
Research Article (10 pages), Article ID 9912265, Volume 2021 (2021)

Research Article

Industrial Internet Federated Learning Driven by IoT Equipment ID and Blockchain

Xu Zhang¹, Haibo Hou¹, Zhao Fang² and Zhiqian Wang²

¹China Academy of Information and Communications Technology, China

²Guangzhou Institute of Internet of Things, China

Correspondence should be addressed to Xu Zhang; zhangxu1@caict.ac.cn

Received 9 July 2021; Revised 12 September 2021; Accepted 20 September 2021; Published 8 November 2021

Academic Editor: Bo Rong

Copyright © 2021 Xu 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.

With the development of Internet of Things (IoT), 5G, and industrial technology, Industrial Internet has become an emerging research field. Due to the industrial specialty, higher requirements are put forward for time delay, safety, and stability of the identification analysis service. The traditional domain name system (DNS) cannot meet the requirements of industrial Internet because of the single form of identification subject and weak awareness of security protection. As a solution, this work applies blockchain and federated learning (FL) to the industrial Internet identification. Blockchain is a decentralized infrastructure widely used in digital encrypted currencies such as Bitcoin, which can make secure data storage and access possible. Federated learning protects terminal personal data privacy and can carry out efficient machine learning among multiple participants. The numerical results justify that our proposed federated learning and blockchain combination lays a strong foundation for the development of future industrial Internet.

1. Introduction

In recent years, countries worldwide have paid more and more attention to the development of Industrial Internet. Industrial Internet is an important cornerstone of the fourth industrial revolution and a key measure to transform old kinetic energy into a new one [1]. With the development of Industrial Internet in the past few years, a single form of equipment and different types of enterprises have been connected by the Industrial Internet, which allows the resources of different links to be organically combined. The Industrial Internet system architecture consists of four aspects: network connection, platform, security system, and identification analysis system [2]. Among them, the network is used to realize the connection of people, machines, and things, and it is the foundation of Industrial Internet. The security system is responsible for providing security protection and guarantee, and the purpose of the platform is to open up operational data and Internet data to integrate resources. The identification analysis system is an important hub for the realization of Industrial Internet.

The traditional domain name system (DNS) resolution service faces serious challenges in terms of subject identification, resolution methods, security, and service quality, and it cannot meet the needs of industrial networks. The main reasons can be summarized as follows: change of subject identification, mass data and ultralow latency requirements, security and privacy protection, fairness, and reciprocity. It is mainly because the blockchain system has the characteristics of calculation, storage, and scalability, while federated learning (FL) has the characteristics of ensuring information security during big data sharing and exchange and protecting the privacy of terminal personal data. This paper uses blockchain to store important information about devices; the model uses the convolutional neural network (CNN) of FL as the baseline. The technologies of blockchain and federated learning are applied to Industrial Internet identification analysis, which enables identification analysis to play a more important and irreplaceable role in the Industrial Internet field.

The chain storage structure of the blockchain can comprehensively record the data generated by Industrial Internet companies in the production and operation process, which

makes the data nontamperable, thereby ensuring the authenticity and credibility of the data. This is also advantageous for Industrial Internet companies to reduce costs and improve efficiency. The privacy protection of blockchain technology are applied due to the advantages of trusted collaboration; it can be deeply integrated with the Industrial Internet in terms of data confirmation, accountability, and transactions, thereby promoting the transformation of industrial production to digital and intelligent [3]. Feng et al. [4] introduced four core technologies of blockchain: decentralization, consensus mechanism, encryption algorithm, and smart contracts. In the blockchain, users jointly create a public ledger for block verification and transaction records [5]. [6]. Blockchain technology has laid a solid foundation for earning trust and created a reliable cooperation mechanism, and it has a wide range of application prospects [7, 8].

Federated learning is a solution for machine learning and artificial intelligence (AI) to face more stringent data management regulations. In the framework of federated learning, the central server saves global data that can be initially shared, and each client saves local data and trains local machine learning and artificial intelligence models based on the local data. Then, according to a certain communication mechanism, client transmits the model parameters and other data to the central server. The central server collects the data uploaded by each client and conducts training to build a global model; each client has the same role and status in the entire federated learning mechanism [9]. Federated learning effectively solves the problem of the client sharing data between two or more data without contributing data, so it solves the problem of data islands to a large extent.

The sensors and IoT devices deployed in the Industrial Internet of Things generate massive amounts of sensor data, and the analysis of sensor data can promote industrial production and manufacturing. When federated learning analyzes and processes massive amounts of sensor data, there is no need for data interaction between devices, so the privacy of local data can be guaranteed. For example, in the process of anomaly detection in the Industrial Internet of Things [10], the privacy of local data can be guaranteed by using federated learning, and there is no need to interact with local data between devices, which can improve the ability to detect abnormal IoT nodes in the process of anomaly detection. At the same time, data-driven cognitive computing (D2C) faces some important bottlenecks in the Industry 4.0 scenario [11]. In order to solve the problem of privacy leakage in cognitive computing, federated learning can be used in cognitive computing in the Industrial Internet of Things to protect data security and prevention of privacy leaks. However, if the central server fails or has a trust issue, then all computing and information security cannot be guaranteed, and a single point of failure will occur [12]. The decentralized distributed data storage structure of the blockchain can remove the trusted central authority, so it can solve the trust problem of the central server in the federated learning, thereby preventing the single point of failure. Therefore, the use of blockchain technology based on federated learning can prevent single-point failure problems, and the verification mechanism of the blockchain can ensure the

authenticity of data while selecting high-quality and credible edge device [11].

The main contributions of this article are summarized as follows:

- (1) It is necessary to include privacy protection in the equipment identification of Industrial Internet. We utilize federated learning to guarantee information security during large data sharing and exchange, as well as safeguard the privacy of terminal personal information
- (2) We propose a novel framework that applies blockchain and federated learning technology to the research of Industrial Internet identification. The performance of the proposed framework is verified by solid numerical results

The rest of this paper is organized as follows. Section 2 presents the related work. Section 3 introduces blockchain technology as well as its application in industrial Internet. Section 4 proposes our solution of blockchain-based federal learning with numerical results, followed by Section 5 to conclude the paper.

2. Related Work

With the rapid development of Internet of Things (IOT), 5G networks, and industrial technology, some new applications such as smart cities, virtual reality, and industrial intelligent production continue to emerge. The number of wearable devices, industrial machines, and different types of sensors has exploded, which indicates that the future network is transforming from a consumer to a production model. The particularity of industrial production requires that industrial networks can perceive environmental information through intelligent means, support a large number of heterogeneous device access, support massive multisource, multimodal data high-speed transmission, and have stronger security, thus providing better service for enterprise production. This has brought huge challenges to the traditional Internet in terms of architecture, security, and performance [13].

Different from the consumer Internet and the traditional IOT, industrial Internet has diverse communication subjects and higher performance requirements, and traditional DNS resolution services cannot meet their needs. In order to meet the characteristics and requirements of industrial Internet, its identification resolution system must follow these principles: support for multisource heterogeneous communication subjects, security guarantee for identification resolution services in complex environments, fair and equal guarantee for participation of multiple organizations, effectiveness guarantee in scenarios with multiple protocols, high concurrency and differentiated requirements, and providing scalability at the protocol level and system level. In response to the principles, considering the calculation, storage, and scalability characteristics of the blockchain system, federated learning has the characteristics of ensuring information security during big data sharing and exchange and

protecting the privacy of terminal personal data. The technology of blockchain and federated learning are applied to industrial Internet identification analysis, which promote the development of identification analysis in the field of industrial Internet.

Blockchain is a new type of technology that has gradually emerged with encrypted digital currency. It adopts a distributed computing model, uses blockchain to store data, and uses cryptographic principles to ensure the security of transmission and access. Data storage is jointly maintained and supervised by Internet users. It has distinctive features such as decentralization, transparency and openness, and unmodifiable data [14]. It uses distributed storage and calculation to ensure that the entire network node has the same rights and obligations, and the data in the system are essentially maintained by the nodes of the entire network. By this means, the blockchain no longer depends on the central processing node to realize the distributed storage, record, and update of data. Therefore, its application is not limited to currency as an asset type, and its application research in various industries is relatively extensive. Chi et al. [15] summarized the existing blockchain technologies in several typical fields, and then gave the main problems and countermeasures in the development process of blockchain in detail, and finally discussed the prospects and forecasts of blockchain. Blockchain is a distributed ledger technology that relies on logical control functions such as smart contracts to evolve into a complete storage system. Changes in its classification methods, service models, and application requirements have spawned to the diversified core technologies. In order to fully understand the blockchain ecosystem, Bao et al. [16] designed a hierarchical blockchain technology architecture and further analyzed the basic principles, technical associations, and research progress of each layer structure of the blockchain. The frontier application directions of blockchain such as the industrial Internet and smart cities are given at the end.

Federated learning based on client-server architecture and distributed machine learning [17] are both used to process distributed data, but there are differences between them in terms of application fields, data attributes, and system composition [18]. Federated learning algorithms can be divided into machine learning-based algorithms and deep learning-based algorithms. Federated learning has the following characteristics.

- (i) Supporting nonindependent and identically distributed data: the federated learning algorithm performs well in nonindependent and identically distributed data. In the actual use of federated learning, the data quality and distribution of the data holder is uncontrollable. The data of the holder cannot be required to meet independent and identical distribution, so the federated learning algorithm needs to support nonindependent and identically distributed data
- (ii) Efficient communication: federated learning needs to consider the system heterogeneity of the data holder to improve communication efficiency and reduce communication loss without losing accuracy or loss

- (iii) Fast convergence: in the process of joint modeling, it is necessary to ensure the convergence of the model and at the same time increase the convergence speed
- (iv) Security and privacy: since data privacy security is an important feature of federated learning, security and privacy are two necessary requirements for federated gradient updates. Security and privacy can be carried out in the aggregation process through encryption and other methods and can also be reflected in the process of stand-alone optimization
- (v) Support complex users: complex users refer to the large number of users and the imbalance or deviation of user data. The federated optimization algorithm needs to have good compatibility ability to deal with this situation

Based on the advantages of the above federated learning algorithm, information security can be guaranteed during big data sharing and exchange, and the privacy of terminal personal data is protected. Furthermore, it can carry out high-efficiency machine learning among multiple participants or multiple computing nodes. Zhang et al. [19] made a comprehensive review of recent research and achievements in federal learning and presented future development trends. First, data islands and privacy protection are described to introduce the background of federated learning, and the connotation and mechanism of federated learning are outlined. Then, typical application cases of data sharing and exchange based on federated learning technology are introduced.

3. Blockchain for Industrial Internet

Currently, mainstream object recognition systems include object identifier (OID) recognition system, Ecode recognition system, and handle recognition system. The OID identification system has simple coding rules, good flexibility, and scalability. Therefore, the OID identification system is adopted to apply the traceability system of industrial equipment. We first propose the overall framework design of the system in this section. At the same time, in order to ensure the security of the system, the key and authentication mechanism and the authority management mechanism are proposed, and then, the traceability scheme inside the blockchain and the traceability scheme outside the blockchain are designed in detail.

3.1. System Security Design

- (1) Key and authentication mechanism: the key and authentication mechanism mainly include the identification of the traceable company's identity information and the distribution of key pairs. It uses cryptographic techniques such as asymmetric encryption, digital signatures, and public key infrastructure (PKI) authentication systems. A traceable enterprise that successfully performs identity verification can obtain a key pair issued by the key

management center. The key pair includes a public key and a private key. When the user calls the smart contract, the verifier will verify the data on the chain. After confirming that the data is legal, the public key will be used to encrypt the data, and then, the traceability information will be recorded in the blockchain. When reading the information in the blockchain, the smart contract can be called to obtain encrypted data, and the corresponding traceability information can be obtained after decryption with the private key

All traceable companies need to perform identity authentication and key distribution operations before joining the system. The detailed steps of key and authentication are as follows:

- (i) The traceability company submits the key pair and digital certificate application to the certification center. During the application process, it is necessary to provide traceable company certification materials, including information such as the company's social credit code and company name
 - (ii) The certification center reviews the applicant's certification materials. After passing the identity authentication, the certification center submits the applicant's information to the key management center and requests the distribution of key pairs. After receiving the request, the key management center generates a key pair, adds the flag field information, and saves it in the secure database. The marked field is the social credit code of the traceable enterprise, which is used to identify the identities of different applicants
 - (iii) The key management center sends the key pair to the applicant and at the same time returns the public key to the certification authority center to generate a digital certificate
 - (iv) The certification center generates a digital certificate based on the applicant's identity information and the public key and returns the digital certificate to the applicant. The certificate adopts the X.509 standard proposed by ITU-T. The key management center will return the key pair, and the certificate authority will return the digital certificate. The applicant can compare them to determine whether the key pair is correct. If there is an error, it needs to reapply
- (2) Key and authentication mechanism: in the industrial equipment traceability system, certain transaction information has a certain degree of confidentiality and can only be accessed by specific users. However, blockchain technology has the characteristics of information transparency and information sharing. Any user of the unrestricted blockchain network can obtain the information in the block, which leads

to the leakage of users' private information. In this regard, two propose the user rights management. Different types of users have different access rights, thereby ensuring the security of user information and preventing the leakage of user privacy information. We assign permissions based on the tasks and needs of users in traceability companies, regulatory agencies, and industrial equipment traceability systems

- (i) Supervision department: it is necessary to supervise all traceability information of industrial equipment. Therefore, the supervision department has the highest access authority in the traceability system and can add, delete, modify, and query traceability information. Adding operations is to call smart contracts to record traceability information on the blockchain. The delete operation will not directly delete the traceability information from the blockchain, but will add the traceability information corresponding to the state of the industrial equipment in the "deleted" state to the block. There are seven industrial equipment statuses in the traceable system, which are transportation, delivery, distribution, sale, return to the factory, and deletion. Similarly, the modification operation also adds traceability information of the state of the industrial equipment to the block. The query operation can query the complete traceability information table of industrial equipment
- (ii) Traceability companies: it provides traceability services general access rights and can add, modify, and query traceability information. Among these operations, the addition and modification operations are the same as the highest access authority. For query operations, general access rights can only query the production, circulation, distribution, and supervision of industrial equipment
- (iii) User: it needs to query the traceability information of the purchased product and has the lowest access authority. Besides, it can only perform query operations that is the same as the traceability company

3.2. Block and Traceability Information Table. Equations should be provided in a text format, rather than as an image. Microsoft Word's equation tool is acceptable. Equations should be numbered consecutively, in round brackets, on the right-hand side of the page. They should be referred to as Equation (1), and so on in the main text.

The block and traceability information is the most basic data structure in the traceability scheme in the blockchain, which mainly includes three parts: block, transaction table, and traceability information table. Blockchain is a chained data structure composed of multiple blocks. A block is a carrier used to store transaction orders. Each transaction order is a piece of traceability information for industrial

equipment. The user calls the smart contract to obtain the traceability information table of the industrial equipment through the unique OID number of the industrial equipment. Due to different access rights, the traceability information tables obtained by the user are also different.

The block and traceability information table is shown in Figure 1, which mainly includes three parts: block, transaction ticket, and traceability information form. In the block part, the first is the block header, which is composed of the timestamp of the block generated by the hash value of the previous block and the hash value of the root of the Merkle tree. Common hash algorithms include SHA1, SHA2, and MD5. In this paper, we use the SHA-256 hash algorithm, whose reliability and security meet the requirements of the traceability system. The second is the block body, which includes a single transaction number and transaction order. In the transaction ticket part, the transaction table is mainly composed of the product identification code, digital abstract, transaction content, timestamp, public key, and digital signature. The part of the traceability information table mainly includes five links. The product identification code is the identification mark of industrial equipment.

3.3. Blockchain Traceability Scheme. In the device traceability system based on the blockchain, the traceability enterprise first needs to perform identity authentication and key distribution. After the operation is completed, each participant in the system is assigned a different authority. When the equipment circulates in the supply chain, traceability companies, and regulators will call smart contracts to record equipment traceability information.

In the traceability solution outside the blockchain, we apply the OID identification system to the equipment traceability system. We store equipment traceability information in the traceability enterprise identification management server through the OID identification information registration mechanism and obtain detailed equipment traceability based on the OID identification analysis mechanism information to achieve equipment traceability. The external information of capacity equipment traceability blockchain management solves the problem of blockchain data explosion.

The caption can also be used to explain any acronyms used in the figure, as well as providing information on scale bar sizes or other information that cannot be included in the figure itself. Plots that show error bars should include in the caption a description of how the error was calculated and the sample size (see Figure 2).

3.4. Experiment Results Analysis. This section mainly tests the actual operating efficiency of the improved blockchain and analyzes the test results to verify the practicability of the proposed traceability system. First, an improved blockchain on a virtual machine is deployed. After that, sending suggestions and querying requests are used to test the system latency and throughput to activate the blockchain network. In the paper, throughput refers to the number of requests processed per unit time, and its unit is Tx/s. The system delay and throughput are obtained after many tests.

Figure 2 shows the system throughput under the proposal requests and the query requests. The system throughput increases at the beginning and then decreases with the increase of the proposal request. When the proposal request reached 4000, the throughput reaches the highest value. This is because the number of requests exceeds the processing capacity of the node, which will cause thread blocking and reduce system performance. At the same time, we can see from Figure 2, regardless of the query request, the system throughput is basically stable at about 350. This is because during the query request process, the blockchain only performs read operations instead of the write operations, which does not occupy system resources. As a result, the system throughput is relatively stable.

4. Blockchain-Based Federated Learning

The application of FL in the industrial Internet of Things (IIoT) is introduced in this section. We propose a FL framework based on device recognition in IIoT, which takes into account communication efficiency and data privacy.

4.1. The Communication Efficiency of the IIoT. There are two communication modes for the IIoT, including wired communication and wireless communication [20]. Recently, wireless communication technology is widely used in various fields, especially in the IIoT because of its flexibility and scalability. In the IIoT communication system, devices in the network are usually in different environments. In addition, there are a large number of devices in the IIoT, which require the communication system to be flexible and expandable. Therefore, it is a natural trend to regard wireless communication technology as the main method of the IIoT. With FL technology, local devices need to iteratively upload gradients to the central server, which introduces enormous communication overhead. In this case, the top- k algorithm is proposed to reduce communication costs.

4.2. The AI Model Based on Device Identification. Traditional machine learning needs to collect data from multiple devices to a central server for training, which only considers the performance of the central server when an AI model is designed. In this paper, we adopt FL to alleviate the data privacy problems. FL is a distributed machine learning. To consider the performance of all devices in the IIoT, especially in the IIoT, there are a large number of smart devices distributed in the network. The storage and calculation performance of these devices are different (the storage performance of mobile phones is worse than that of notebooks). Therefore, it is essential to consider devices with poor performance to ensure that each device can work successfully. In this paper, the device identification is stored through blockchain technology.

The AI model is designed based on device identification. In this paper, the AI model changes the depth of the convolutional layer according to the performance of the device in the network. We can increase the depth of convolutional layers when devices have superior performance. It is undeniable that the performance of the AI model is better with the

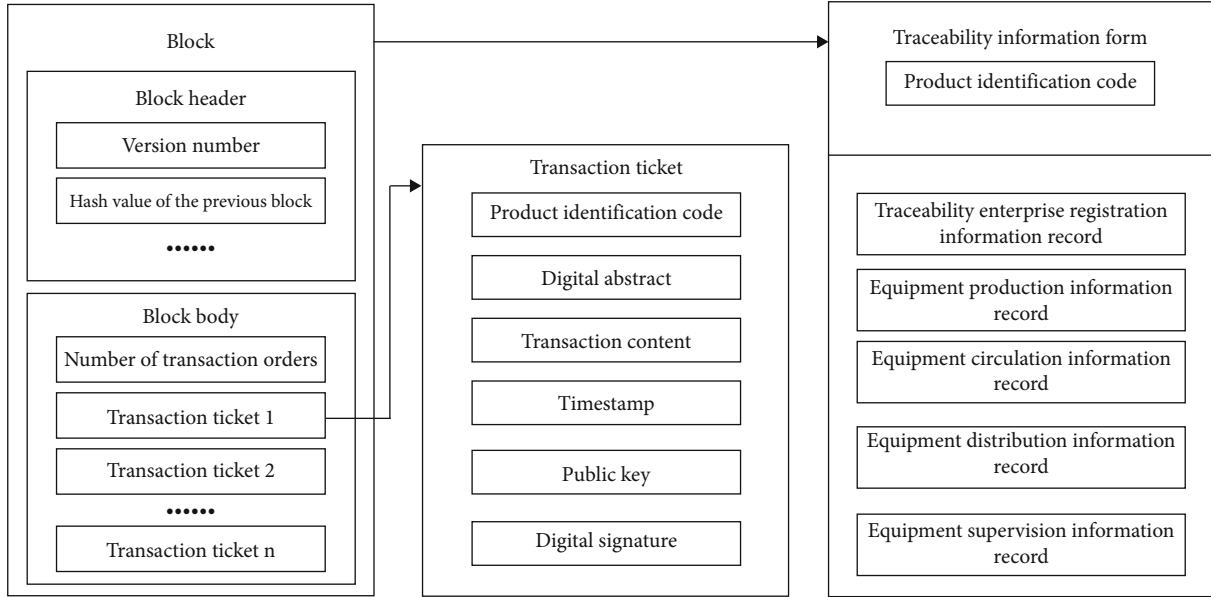


FIGURE 1: Block and traceability information table.

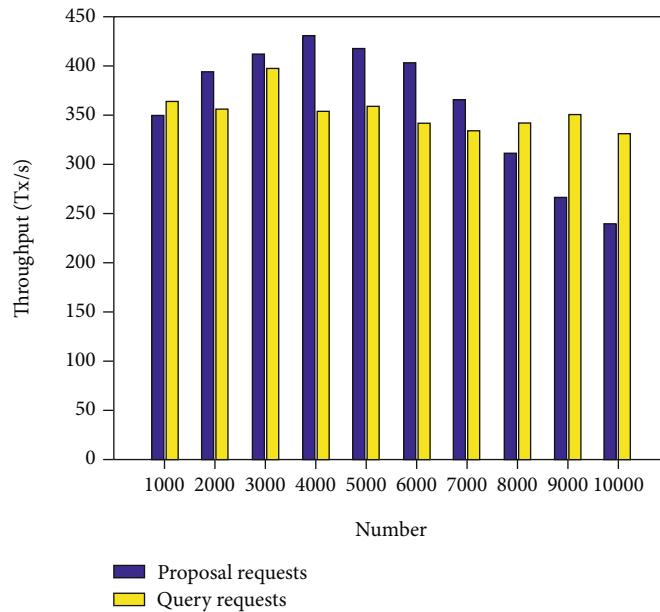


FIGURE 2: System throughput in case of proposal request and query requests.

increment of convolutional layers. This proposed framework designs the AI model based on the performance of the device, which makes a trade-off between the complexity of the network and the performance of the model.

4.3. *Proposed Framework Based on FL.* Traditional machine learning aggregates data from different devices to a central server, which may lead to data leakage and privacy infringement. Users are unwilling to share data, which leads to the problem of data isolated island. FL has been studied by many researchers in order to alleviate these problems. FL realizes data sharing in the IIoT and protects data privacy to a certain extent. However, the application of FL technology

to the IIoT faces many challenges. For example, devices in wireless communication networks upload local gradients or models, which will bring a lot of communication overhead. It is essential to elaborate methods to reduce communication overhead.

Besides, the attacker can ratiocinate the private data from the gradient or model uploaded by local devices [21, 22]. To alleviate the problem of privacy protection in FL, many methods have been proposed, which are mainly divided into encryption technology [23] and differential privacy methods [24, 25]. Methods based on cryptography protect data privacy at the cost of decreasing communication efficiency. In this paper, we propose an algorithm based

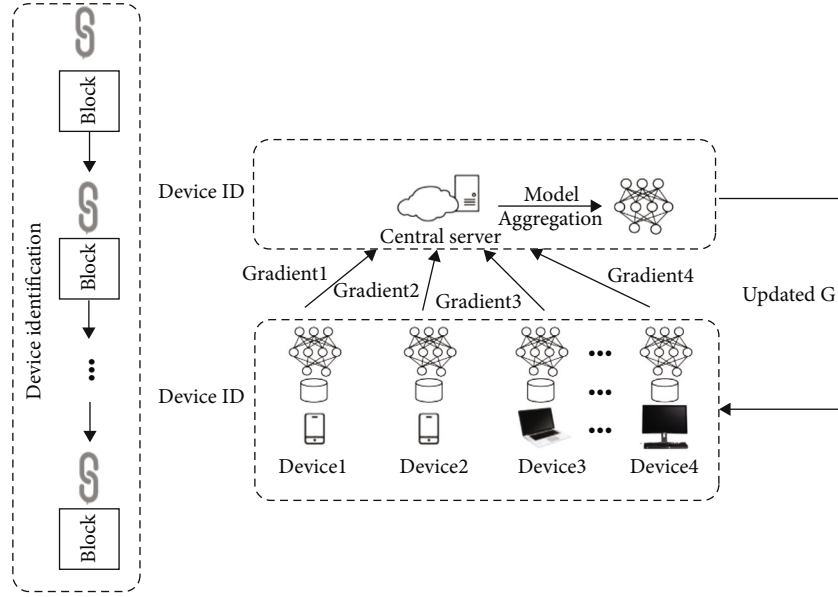


FIGURE 3: The proposed framework based on FL integrated with blockchain.

on differential privacy technology to alleviate the problem of data privacy.

The proposed framework based on FL integrated with blockchain is shown in Figure 3. It describes the system model applied to the industrial Internet scenario, where the underlying device with different performance is trained based on the local dataset. The training model uses the CNN model, and then, the local equipment uploads the model parameters to the central server for aggregation. After aggregation, the new model parameters are sent to each device. Among them, the blockchain can remove the trusted central authority. Therefore, it can solve the trust problem of the central server in the federated learning, preventing the single point of failure in the federated learning. In order to improve the communication efficiency, the proposed method employs a top- k algorithm when the gradient is uploaded to the central server. Besides, in terms of data privacy, we propose an improved differential privacy technology to realize data sharing. The detailed process of FL is exhibited as follows.

Step 1: local model train. The local device trains the AI model iteratively based on the local database to achieve a linear optimization problem locally. ω_i^l, γ represent the local model and learning rate of the device in the i th iteration, respectively, and $f(\bullet)$ is the loss function.

$$\omega_i^l = \omega_i^{l-1} - \gamma \nabla f(\omega_i^l). \quad (1)$$

Step 2: local gradient preprocess. The local device trains the AI model based on the local database. To reduce the communication overhead, the proposed framework employs the top- k algorithm. Device D_i calculates the absolute value of the parameters in epoch t . Then, these values are sorted in positive order. We choose the first k values for global aggregation. This paper uses Gaussian mechanism to add

noise to the local gradient, which protects the data privacy of users. The gradients in epoch t can be expressed as

$$g_t(x) \leftarrow OL(\omega). \quad (2)$$

The process of gradient clipping is as

$$g_t(x)' = \frac{g_t(x)}{\max(1, \|g_t(x)\|_2/S)}. \quad (3)$$

Step 3: local gradient calculation. The selected k gradients utilize differential privacy for noise disturbance, where S denotes the L_2 -norm threshold of gradient clipping. Note that S is the global sensitivity set in advance. The procedure ensures that the L_2 -norm of the gradients of the local device is within the range of S . Then, Gaussian noise is added to the gradients which introduces randomness. The gradients adding Gaussian noise are as

$$g_t(x)'' = \frac{1}{b} \sum (g_t(x)' + N(0, \sigma^2)), \quad (4)$$

where b presents batch size and $N(0, \sigma^2)$ represents Gaussiandistribution. The variance σ must meet

$$\sigma = \frac{S}{e} \sqrt{2 * \ln(1.25/\delta)}, \quad (5)$$

where δ presents slack factor, s presents privacy budget, and s is set differently based on the data size of the local dataset. Besides, s is modified based on the loss of the model.

Step 4: local model update. Randomness is introduced due to the noise. The local model is updated by

$$\omega_{t+1} = \omega_t - \eta * g_t(x) \quad (6)$$

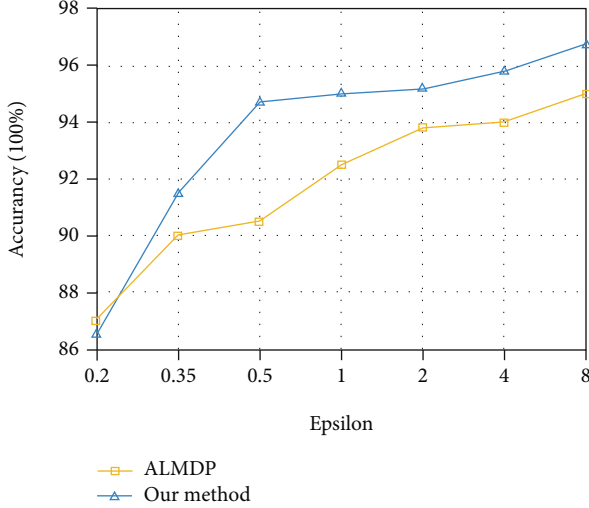


FIGURE 4: Accuracy of the proposed framework under different epsilons.

Step 5: local gradient upload. Devices in the industrial Internet upload local gradients to the central server.

Step 6: model aggregate. The central server aggregates the gradients from devices in the IIoT. Then, the server sends the new model to all devices. The global model of epoch $t + 1$ is shown as

$$\omega_{t+1} = \omega_t + \frac{1}{m} \left(\sum_{k=1}^m \Delta \omega_{t+1}^k \right), \Delta \omega_{t+1}^k = \omega_{t+1}^k - \omega_t^k. \quad (7)$$

4.4. Experiment Result Analysis. To evaluate the performance of the proposed method, a well-known digital classification dataset MNIST is employed. The AI model is designed based on the performance of the devices in the IIoT. Users can get the performance by the device identifier stored in the blockchain. There are 40 devices (15 mobile phones and 25 computers) in the wireless work to verify the proposed framework. We employ the CNN model as the learning model. Mobile phones are given in the network. The CNN model is designed with 2 convolutional layers and 2 fully connected layers. The kernel size is set as 3×3 . The gradient clipping threshold S is set to 0.01. The AI model employs the maximum pooling and dropout to alleviate overfitting. Classification accuracy is employed as the evaluation criteria. To realize differential privacy protection, this paper adds Gaussian noise to the gradients in the process of gradient descent of backpropagation. As is shown in Figure 4, we compare the performance of our method with the ALMDP [24] algorithm. The result shows that our method is better. The accuracy of the model can reach 97.6%, which increases with the enlargement of s .

The performance of FL without differential privacy is compared with our proposed framework. As shown in Figure 5, the accuracy of FL without differential privacy is better than our method before epoch 12.

However, the accuracy of FL without differential privacy decreases rapidly in epoch 12 due to malicious modification

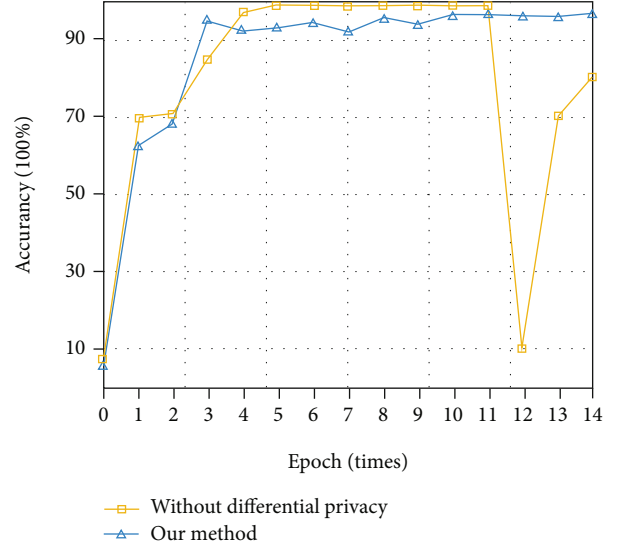


FIGURE 5: Performance comparisons on our method and FL without differential privacy.

of one device. Even if the performance of our method is worse, our method makes a trade-off between performance and data privacy.

5. Conclusions

The industrial Internet has received full attention from domestic and international researchers. As an important infrastructure of the industrial Internet, identification resolution technology is a link that must be overcome. The network interconnection of industrial control systems, the intercommunication of industrial data, and system security are issues that require great attention during the development of the industrial Internet. Blockchain technology has the characteristics of decentralization, immutability, and low cost, which can solve the pain points in the development process. Federated learning, as the basic theory of large-scale collaboration in the next generation of artificial intelligence, provides effective solutions to key issues such as small data and privacy in the current development of artificial intelligence, which can further promote the development of the industrial Internet. The numerical results justify that the federated learning and blockchain technology for industrial Internet identification proposed in this paper are practical and effective, which opens up a new research direction for the development of industrial Internet identification.

Data Availability

This is private company data. For confidentiality reasons, all data are not available to the public.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Ministry of Industry and Information Technology of China, 2019 Industrial Internet Innovation and Development Project Device ID Resolution Service Capability Test and Verification Platform Project (20200151).

References

- [1] J. Li, F. R. Yu, G. Deng, C. Luo, Z. Ming, and Q. Yan, "Industrial internet: a survey on the enabling technologies, applications, and challenges," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1504–1526, 2017.
- [2] E. P. Yadav, E. A. Mittal, and H. Yadav, "IoT: challenges and issues in Indian perspective," in *2018 3rd International Conference On Internet of Things: Smart Innovation and Usages (IoT-SIU)*, pp. 1–5, Bhimtal, India, 2018.
- [3] Y. Cheng and H. Shaoqin, "Research on blockchain technology in cryptographic exploration," in *2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE)*, pp. 120–123, Bangkok, Thailand, 2020.
- [4] J. Feng, Y. Wang, J. Wang, and F. Ren, "Blockchain-based data management and edge-assisted trusted cloaking area construction for location privacy protection in vehicular networks," *IEEE Internet of Things Journal*, vol. 8, no. 4, pp. 2087–2101, 2021.
- [5] A. Kuzmin and E. Znak, "Blockchain-base structures for a secure and operate network of semi-autonomous unmanned aerial vehicles," in *2018 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*, pp. 32–37, Singapore, 2018.
- [6] H. Yang, H. Cha, and Y. Song, "Secure identifier management based on blockchain technology in NDN environment," *IEEE Access*, vol. 7, pp. 6262–6268, 2019.
- [7] A. Irshad, M. Usman, S. Ashraf Chaudhry, H. Naqvi, and M. Shafiq, "A provably secure and efficient authenticated key agreement scheme for energy internet-based vehicle-to-grid technology framework," *IEEE Transactions on Industry Applications*, vol. 56, no. 4, pp. 1–4435, 2020.
- [8] A. Dua, N. Kumar, A. K. Das, and W. Susilo, "Secure message communication protocol among vehicles in smart city," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 4359–4373, 2018.
- [9] S. Niknam, H. S. Dhillon, and J. H. Reed, "Federated learning for wireless communications: motivation, opportunities, and challenges," *IEEE Communications Magazine*, vol. 58, no. 6, pp. 46–51, 2020.
- [10] Y. Liu, N. Kumar, Z. Xiong, W. Y. B. Lim, J. Kang, and D. Niyato, "Communication-efficient federated learning for anomaly detection in industrial internet of things," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, pp. 1–6, Taipei, Taiwan, 2020.
- [11] Y. Qu, S. R. Pokhrel, S. Garg, L. Gao, and Y. Xiang, "A blockchain-federated learning framework for cognitive computing in industry 4.0 networks," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 4, pp. 2964–2973, 2021.
- [12] J. Passerat-Palmbach, T. Farnan, M. McCoy et al., "Blockchain-orchestrated machine learning for privacy preserving federated learning in electronic health data," in *2020 IEEE International Conference on Blockchain (Blockchain)*, pp. 550–555, Rhodes, Greece, 2020.
- [13] F. Li, A. Yang, H. Chen et al., "Towards industrial internet of things in steel manufacturing: a multiple-factor-based detection system of longitudinal surface cracks," in *2020 IEEE International Conference on Big Data (Big Data)*, pp. 4627–4635, Atlanta, GA, USA, 2020.
- [14] D. Gräf, M. Friedlein, C. Gänßmantel, J. Franke, and N. Ischdonat, "New concept for the integration of additive manufactured mechanical and mechatronic components in aircraft interior systems," in *2020 Advances in Science and Engineering Technology International Conferences (ASET)*, pp. 1–5, Dubai, United Arab Emirates, 2020.
- [15] C. Chi, D. Han, Q. Zhang et al., "Research on distributed new energy spot trading method based on blockchain Technology," in *2020 Chinese Automation Congress (CAC)*, pp. 275–278, Shanghai, China, 2020.
- [16] Z. Bao, Q. Wang, W. Shi, L. Wang, H. Lei, and B. Chen, "When blockchain meets SGX: an overview, challenges, and open issues," *IEEE Access*, vol. 8, pp. 170404–170420, 2020.
- [17] R. Wakayama, R. Murata, A. Kimura, T. Yamashita, Y. Yamauchi, and H. Fujiyoshi, "Distributed forests for MapReduce-based machine learning," in *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pp. 276–280, Kuala Lumpur, Malaysia, 2015.
- [18] M. Aledhari, R. Razzak, R. M. Parizi, and F. Saeed, "Federated learning: a survey on enabling technologies, protocols, and applications," *IEEE Access*, vol. 8, pp. 140699–140725, 2020.
- [19] W. Zhang, X. Wang, P. Zhou, W. Wu, and X. Zhang, "Client selection for federated learning with non-IID data in mobile edge computing," *IEEE Access*, vol. 9, pp. 24462–24474, 2021.
- [20] Y. Li, Y. Ma, Z. Yin, A. Gu, and F. Xu, "A communication model to enhance wireless industrial networks based on time-sensitive networks," in *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*, pp. 363–367, Chengdu, China, 2020.
- [21] H. Yang and Y. Zhou, "AIC-GAN: an auxiliary information classification GAN for learning deep models," in *2020 Chinese Automation Congress (CAC)*, pp. 6106–6111, Shanghai, China, 2020.
- [22] L. Melis, C. Song, E. De Cristofaro, and V. Shmatikov, "Exploiting unintended feature leakage in collaborative learning," in *2019 IEEE Symposium on Security and Privacy (SP)*, pp. 691–706, San Francisco, CA, USA, 2019.
- [23] L. T. Phong, Y. Aono, T. Hayashi, L. Wang, and S. Moriai, "Privacy-preserving deep learning via additively homomorphic encryption," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 5, pp. 1333–1345, 2018.
- [24] N. Phan, X. Wu, H. Hu, and D. Dou, "Adaptive Laplace mechanism: differential privacy preservation in deep learning," in *2017 IEEE International Conference on Data Mining (ICDM)*, pp. 385–394, New Orleans, LA, USA, 2017.
- [25] P. C. Mahawaga Arachchige, P. Bertok, I. Khalil, D. Liu, S. Camtepe, and M. Atiquzzaman, "Local differential privacy for deep learning," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 5827–5842, 2020.

Research Article

Informatization Construction of Digital Assets in Smart Cities

Wenze Ning  and Mei Lu

College of Management, Xi'an University of Architecture and Technology, Xian, 710055 Shanxi, China

Correspondence should be addressed to Wenze Ning; nwz@xauat.edu.cn

Received 21 June 2021; Revised 18 September 2021; Accepted 30 September 2021; Published 22 October 2021

Academic Editor: Peng Yu

Copyright © 2021 Wenze Ning and Mei Lu. 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 popularization of the Internet of Things (IoT), smart cities driven by IoT technology have become a matter of concern. This study mainly investigates the construction of digital asset based on the data collected from government departments, enterprises and institutions. Particularly digital asset personnel should not only get familiar with the process of smart city information management, but also be able to identify the risks in digital assets. Strict accountability mechanism can ensure the data privacy, dispel the worries of data providers, and help weaken data barriers. Supported by comprehensive numerical results, we reach a conclusion that our proposed digital asset construction scheme has practical significance in many aspects of smart city development and will help the regional Informatization considerably.

1. Introduction

Internet of Things (IoT) refers to a number of smart physical devices that are connected to each other through the Internet. Through the information transmission via the Internet, smart devices complete the communication and interaction behaviors between each other and with users. With the continuous progress of wireless communications and Internet, IoT technology is also developing rapidly. Internet of Things has already played its role in a variety of critical applications such as remote healthcare, environment monitoring, smart grid, etc. However, raw data from IoT device is often not easily understood, and thus requires conversion to human-readable information. With those pre-processed data, people enter the era of information explosion and sink in the ocean of information.

IoT empowered Smart cities have become the evolution trend of a new type of town. This tendency serves as not only a catalyst for the development of a new kind of community, but also as a strong motivator and a clear vision. Due to the new kind of town's greenness and outstanding environmental and energy-saving features, the smart city built on IoT is the top option. The smart city powered by IoT may help accelerate the development of new cities. The networked

smart city solves many issues associated with urbanization's growth and is an efficient solution for new urbanization.

With the popularization of digital devices and social network applications, the explosive growth of multimedia big data has brought many challenges for users to securely obtain them in various application scenarios. Zhu C first reviewed the latest work of multimedia big data, where the key issues are identified to ensure the success of secure multimedia big data in trust-assisted sensor cloud (TASC). He proposed two types of TASC: TASC-S (TASC with a single trust value threshold) and TASC-M (TASC with multiple trust value thresholds). Although the throughput of TASC-M in his research may fluctuate with the same trust value threshold, the research process lacks data [1]. Hashem I A T proposes a future business model with the goal of managing big data in smart cities. And identified and discussed business and technical research challenges. By focusing on how big data fundamentally changes the urban population at different levels, the vision of supporting big data analysis for smart cities was discussed. This research can serve as a benchmark for researchers and industries to advance and develop smart cities in the context of big data [2]. Li Y studied the current situation of data over-collection and studied some of the most common cases of data over-collection. By

proposing a mobile cloud framework, by putting all user data in the cloud, the security of user data can be greatly improved [3]. Menouar H proposed that there is no smart city without a reliable and efficient transportation system. The realization of next-generation intelligent transportation system (ITS) relies on the effective integration of connected cars and autonomous vehicles. It is also worth noting that ITS that supports drones can be used in the next-generation smart cities. Potential and challenges [4].

Smart city operation center must comply with relevant data management and control standards for high-quality collection when gather data directly by itself. Alternatively the smart city operation center may also need to indirectly collect data from government departments, enterprises and institutions. This requires the centre to negotiate with the data supplier and establish a generally acceptable agreement with the assistance of high-level management agency. Data creation has a direct impact on the quality of subsequent data and the market acceptability of data products. The department of digital asset management may use appropriate tools and technology to create market-ready data products while maintaining data quality and security. Manager of digital asset plays a critical role in the strategic planning of information management in smart cities. Thus, the novelty of this work is that it integrates the digital asset of smart cities with the process of information building to examine the relationship between the two and better support the process of contemporary information construction.

2. Smart City Digital Assets

2.1. Smart City. Recent IoT revolution is growing rapidly with the development of communication technology. Smart cities are the main scenarios for IoT applications, allowing it to contribute to the improvement and progress of cities. The implementation of IoT brings great complexity, and its application requires great efforts to achieve. The deployment of IoT is critical to building smart cities due to communication and data management. With the development and implementation of IoT in cities, many problems will arise. These issues can be categorized by network type, flexibility and scalability, heterogeneity, and end user participation. The need for high-speed connectivity to support connections between different IoT devices is huge. Another major requirement is data storage for all connected devices. To complement existing communication and data management solutions, a large number of software applications and hardware devices are needed to manage and operate user-facing applications. Sustainability and efficiency of software applications and hardware devices are difficult to achieve. Smart cities can connect healthcare. A central connection manager with pollution control, water management, power management, traffic control, security and privacy control devices/tools. It is difficult for a single central management facility to manage and deliver services using existing technical solutions.

The main road of industrial development is definitely the construction of smart city based on IoT. However, in the future, there is no technical support for the implementation

of information architecture and big data in other fields [5, 6]. The urban management pattern is large. From the beginning, it has a high positioning and a long-term vision. It fully considers the current situation and future development needs of the digital construction of urban management. In the system construction, sufficient access ports are reserved for future urban management content refinement and urban management content expansion. At the same time, it pays attention to the construction and sharing of resources within the city. In terms of components and events, the urban management content more comprehensively refines the urban management content at all levels of the city. With the continuous improvement of the digital operation of urban management in the main urban area, it gradually advances to the sub-city area and expands the management scope of urban management digitalization. Reduce the urban management gap between the main urban area and the sub-urban area in the city, and comprehensively improve the digital level of urban management [7]. The overall layout of the smart city is shown in Figure 1.

Adjust urban buildings and public facilities, optimize alternatives, implement 3D spatial analysis of urban buildings, and carry out digitalization and automation of building approval management $W(x, z)$ [8, 9].

$$W(x, z) = \sum_{k=1}^m \sum_{i=1}^n a_{ki} x_{ki} z_i + \sum_{i=1}^n \sum_{j=1}^l b_{ij} x_{ij} z_i + \sum_{k=1}^m \sum_{i=1}^n v_{ki} z_i + \sum_{k=1}^m \sum_{i=1}^n z_i F \quad (1)$$

Among them, x and z are corresponding digital asset information management services [10]. As an important direction of future GIS technology development, 3D GIS uses 3D space coordinates to simulate and visualize the real world [11, 12]. Index eigenvalue relative member matrix R, standard eigenvalue relative member matrix S and relative member matrix M of all levels.

$$M = u_{hj} \left\{ \sum_{i=1}^m [w_{ij} (r_{ij} - s_{ih})]^p \right\}^{\frac{1}{p}} \quad (2)$$

The expected information entropy required to classify the tuple of the data set T according to the attribute A is [13]:

$$T = m(S, t) \frac{f(S, t)}{f(t)} \left[1 - P_c \frac{\delta(S)}{L-1} - O(S) P_m \right] \quad (3)$$

Define the length ∂ , the string length is L [14, 15].

$$\lambda_i^3 = \frac{\partial E}{\partial O_i^3} = \sum_{i=1}^C \frac{\partial E}{\partial O_i^4} \times \frac{\partial O_i^4}{\partial O_i^3} = \sum_{i=1}^C \frac{\partial E}{\partial O_i^4} \times W_{ij} \frac{\sum_{k=1}^{\delta} O_k^3 - O_j^3}{\sum_{k=1}^{\delta} O_k^3} \quad (4)$$

Here, λ_i^3 is the number of rules. The information gain after dividing the data by attribute A is:

FIGURE 1: The overall layout of a smart city(<http://alturl.com/9go98>).

$$P = \frac{\prod_{m=1}^l M_{pm}(z(t))}{\sum_{p=1}^r \prod_{m=1}^l M_{pm}(z(t))} \quad (5)$$

r represents the number of defined gains. There are some unquantifiable indicators in the smart city information security risk assessment index system [16].

$$L = a_1 * \ln \left\{ 1 + \frac{T_e(T_u) - \max(x(K_{10} + K_1))v_1}{1 + T_e(T_u) - \max(x(R_1/K_{10} + K_1))} \right\} \quad (6)$$

Among them, L is the expected value. When the evaluation index attribute is benefit type, the standardization of L is described as [17]:

$$L_{mn} = a_{mn} * \ln \left\{ 1 + \frac{T_e - \min(x(K_{10} + K_1))v_1}{1 + T_e(T_u) - \max(x(R_1/K_{10} + K_1))} \right\} \quad (7)$$

2.2. Asset Informationization. In the digital construction of city management, make full use of the existing platform construction and information resources, do a good job of horizontal resource integration while doing vertical integration, and connect the platforms of other city-level city management related departments and the platforms of subordinate district-level departments. Effectively avoid duplication of construction to achieve resource sharing. In the newly constructed digital platform, high-level configuration and reserved interfaces can be avoided, which can avoid the inability to expand in later construction. It can be found that by integrating the resources of the original urban management related departments and deploying new data resources at a high level, the urban management digital system between the city-level departments and the urban two-level departments can be interconnected, resource-sharing, keep pace, and improve the processing capacity and resource allocation of urban management digitization [18, 19].

$$p = \frac{2}{1 + e^{(-2x)}} - 1 \quad (8)$$

The output layer uses the Log-Sigmoid excitation function, so that the model output value range is in the interval $[0, 1]$ [20]:

$$y = \frac{1}{1 + e^{(-p)}} \quad (9)$$

The output value of the smart city model is denormalized as follows [21, 22]:

$$y_{ij} = y_{\max} - \frac{0.8 - y_{ij}}{0.6} (y_{\max} - y_{\min}) \quad (10)$$

In the formula, the range of y_{ij} is in the interval of $[0.2, 0.8]$. Standardized collection of original indicator data p -dimensional random vector [23]:

$$x = (x_1, x_2, \dots, x_p) \quad (11)$$

The data is standardized and transformed, as shown below [24].

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, p \quad (12)$$

$$\bar{x}_j = \frac{\sum_{i=1}^n x_{ij}}{n}$$

Find the correlation coefficient matrix for the standardized matrix Z , as shown below [25].

$$r_{ij} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_1 - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_1 - \bar{y})^2}} \quad (13)$$

Solve the p characteristic roots of the correlation coefficient matrix, arrange the characteristic roots from large to small, and obtain the orthogonal eigenvectors corresponding to the characteristic roots.

$$\sum_{j=1}^m \lambda_{j'} \sum_{j=1}^p \lambda_j \geq 0.85 \quad (14)$$

3. Digital Asset Construction for Smart Cities

3.1. Establish a Digital Asset Information Management Framework. The digital asset information management framework consists of five sub-modules: digital asset information management, data circulation, value-added services and data privacy protection, data right confirmation and traceability, and digital asset value evaluation.

3.1.1. Digital Asset Information Management

(1) Data Acquisition. Many organizations have established their own databases and information systems in the process of informatization, which leads to differences in the data structures collected by smart city operation centers. These data with different structures, poor quality, and even incomplete data need to be cleaned. In the process of data cleaning, filtering out junk data and correcting data that does not meet the corresponding rules can improve the quality of the data and reduce the problems that may arise in the subsequent digital asset information management work.

(2) Data Circulation. The process of data from the provider to the user is data circulation. From the perspective of smart city data circulation, the amount of data generated by the smart city operation center itself is very small, and most of its data comes from governments, enterprises, social groups, and individuals. The operation center integrates and develops the data collected from these channels, and provides it to government departments, enterprises, social groups and individuals within the scope permitted by relevant regulations. The significance of smart city data circulation is to improve the ability of data users to obtain data resources, and its purpose is to realize the value-added utilization of data through the synchronization of data circulation.

(3) Value-Added Services. To realize the value of digital assets in smart cities, value-added services are the key. In the value-added business of digital assets, data products or data services can generate value through the use of data users. Value-added services achieve the purpose of realizing their potential value by providing additional services or products to data users.

The smart city operation center collects various types of data, and has built a big data analysis platform, so it can provide diversified value-added services. It can not only provide simple data query and download services, but also provide differentiated data products or services as needed, such as customized data visualization analysis or data processing

services. The Smart City Operation Center of W City cooperated with the City Traffic Police Brigade to open a reminder of vehicle violations. During this epidemic, the Smart City Operation Center of W City reported the city's epidemic prevention and control situation in a timely manner to the people of the city through text messages, which produced good social benefits.

(4) Valuation of Digital Assets. When evaluating the value, while taking into account the economic benefits of smart city digital assets, the social benefits of their digital assets must also be considered. The reason for evaluating the value of digital assets is to be able to quantify the value of digital assets. The smart city operation center can classify and manage it based on the quantified digital assets. For example, for data with no value or high value, reduce the collection without violating relevant regulations, and focus on high-value data. In this case, the Smart City Operation Center of W City needs to establish a multi-dimensional value evaluation system based on government benefits, social benefits, and economic benefits.

(5) Data Confirmation and Traceability. Data confirmation mainly refers to clarifying the ownership of data. The property rights here include ownership, possession, control, use, profit and disposal. The concept of data traceability comes from the data life cycle theory, which refers to the evolution of the entire life cycle of traceable data from generation, to use and transaction, to maintenance and destruction. On the basis of data confirmation, the smart city operation center can consider establishing a data traceability system, through blood relationship tracking analysis and other technologies, to manage the whole life cycle of data. Smart city digital asset information management agencies can control the entire process of data from generation to use to destruction, preventing data from being misused or leaked. The organizational structure of digital asset information management is shown in Figure 2.

3.2. Organizational Structure of Digital Asset Information Management. The Smart City Operation Center of W City establishes a management model in which a single department takes the lead and multiple departments coordinate and cooperate. Under the management of the smart city operation center, a separate digital asset information management department can be set up. The management department can conduct professional management of smart city digital assets, be responsible for the daily maintenance and management of digital assets, make overall plans for various expenses in the development or utilization of digital assets, and promote the circulation and value realization of smart city digital assets. Manage the management system and rights protection work exclusively. At the same time, other business departments of the Smart City Operation Center in W City should be separated and assisted in management to promote the effective management of digital assets informatization, and finally maximize the value of the digital assets of the Smart City Operation Center in W City. Builders of smart cities should place a premium on this.

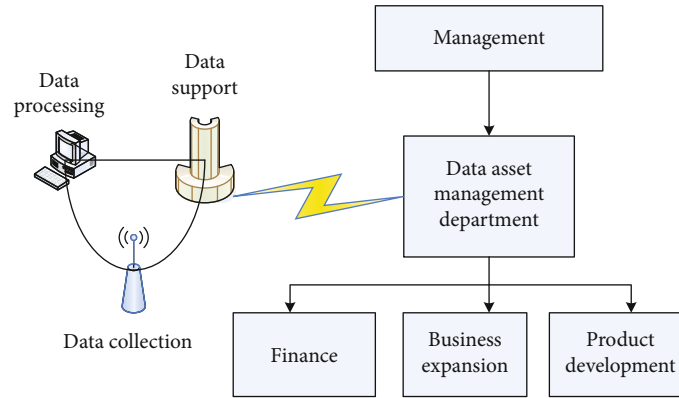


FIGURE 2: Organizational structure of digital asset information management.

By using the city's digital assets as a focal point and value point, it can be foreseen that the future smart city will be more dazzling in terms of culture and tourism. The exhibition of smart cities will emphasize the city's extensive digital assets, rather than burdensome technological solutions.

The management tasks of digital assets are arduous, and there is less experience for special posts. Although the smart city operation center can establish an independent digital asset information management department, it is far from enough to rely on other departments to assist in the management of digital asset identification and value realization. The Digital Asset Information Management Department should be able to independently carry out management work, and consider other departments for assistance when necessary.

3.3. Responsibility Mechanism for Informatization Management of Digital Assets. Through the establishment of a personnel accountability mechanism, all parties involved in the information management of smart city digital assets can clarify their rights and responsibilities, and the rights and responsibilities of various roles in management activities can be refined and ensure the efficient implementation of digital asset informatization management of smart city.

3.3.1. Data Provision. The foundation of digital assets is data. If there is a problem with the basic data provided, the subsequent digital asset information management will be difficult to achieve. Therefore, the data provider is the foundation of the entire digital asset information management. Data providers need to provide corresponding data in accordance with the collection standards of smart city data. When the smart city operation center actively collects data independently, it must also comply with relevant data management and control standards and collect high-standard and high-quality data. On the other hand, smart city operation centers need to indirectly collect data from government departments and enterprises and institutions, and cannot restrict data providers through the agency's responsibility mechanism. This requires the smart city operation center to communicate and negotiate with the data provider, and with the help of the higher-level management organization, reach a consensus on a generally accepted preciousness mechanism.

3.3.2. Data Development. Data development directly determines the quality of the later data and the market acceptance of the demand for data products, whether the value of digital assets in smart cities can be realized, and whether the potential economic benefits of digital assets can be dug out, data development has played a big role. At the same time, the quality and security of the data are also directly responsible for the data developers. The developers of the smart city digital asset information management department can use relevant tools and technologies to develop data products that meet the needs of the market on the basis of ensuring data quality and safety.

3.3.3. Digital Asset Management. Digital asset managers play an overall role in the overall planning of digital asset in smart cities. They must not only know well the process of smart city informatization management, but also be able to identify the risks and timely discover and decisively take appropriate measures for problem fixing. At the same time, they need to have strong communication and collaboration skills and can meet the work requirements of communicating and coordinating with all parties. Through the responsibility recognition mechanism, on the one hand, the responsibility awareness of smart city data providers can be improved, and data providers can be urged to provide qualified data in accordance with the responsibility requirements. On the other hand, a strict responsibility recognition mechanism can ensure the privacy of data providers' data, alleviate the worries of some data providers and help remove data barriers.

4. Numerical Results and Discussion

4.1. Existing Problems in Digital Asset Information Management for Smart Cities. Data operation management statistics are shown in Table 1. Judging from the survey results, most of the respondents do not think that the data operation management of smart cities is doing well. In data life cycle management, nearly 80% of the respondents believe that the performance of smart cities in data life cycle management needs to be improved and general. The data storage of smart cities generally adopts the method of government purchase, and there is no need to consider the cost and profit. For the stored data, there is no need to maximize

TABLE 1: Data Operation Management.

Data operation management	Difference	Needs improvement	General	Good	Excellent
Data life cycle	6.19%	44.62%	39.90%	9.30%	—
Data security management	9.28%	36.42%	34.20%	16.20%	4.10%
Master digital management	43.36%	46.39%	8.25%	2.06%	—

TABLE 2: Data integration and sharing statistics.

Data operation management	Difference	Needs improvement	General	Good	Excellent
Data life cycle	9.28%	59.82%	26.80%	3.10%	1.00%
Data security management	17.53%	42.27%	37.15%	2.06%	1.00%
Master digital management	18.56%	45.36%	35.08%	1.00%	—

TABLE 3: Problems in the information management of digital assets in smart cities.

Problems with digital asset management	W	Y	L	J	Total score
Failure to fully identify and determine	7	9	7	9	32
Failure to establish an effective asset management framework	8	10	7	8	33
Missing asset management organizational structure	8	6	7	8	29
The asset value is not reasonably assessed	8	7	9	8	32
Assets lack operational value-added and circulation	7	8	7	8	30
Barrier phenomenon	8	9	9	8	34

the value. After the data is collected, further application analysis and archiving are not very active. In order to avoid unnecessary troubles, some junk data is not destroyed, and it is still stored in the database. Relatively speaking, respondents have a high evaluation of data security management in smart cities. This is due to the fact that the smart city operation center is also responsible for the construction of government e-government affairs. In order to reduce the risk of government data leakage, the management of smart cities has to increase the construction of data security management. As for the main digital management, respondents generally believe that the digital asset information management of smart cities is relatively poor in this regard. The existing smart city technology framework and solutions also do not reflect the main digital management content.

Data integration and sharing statistics are shown in Table 2. From the survey results, the respondents generally believe that the performance of smart city digital asset information management in data integration and sharing is a need for improvement and a general level. In theory, smart city data should cover all aspects of city operation. However, the reality is that data barriers exist everywhere. As for data integration, most of the respondents believe that the work of connecting, cleaning, converging and integrating smart city data is not in place. When researching on the spot, some research institutions found that the data of some smart city operation centers is only passively stored. This phenomenon is especially seen in some unstructured data. Most of the respondents believe that the sharing of smart city data is not performing well. In the investigation and research on the smart city digital asset link, it is found that the management of the smart city operation center has a negative atti-

tude towards data sharing. It is emphasized that smart cities are not just about implementing next-generation information technologies like the Internet of Things and cloud computing; they are also about establishing a sustainable urban innovation ecosystem characterized by user innovation, open innovation, and popularization. The key data sources of smart city operations are rich, and some of the data belong to government big data. Involving some government data, the management has no legal and regulatory basis for whether these data can be shared. In order to avoid more responsibilities, it is natural not to actively promote the sharing of data. On the whole, in terms of data integration and sharing, smart city digital asset information management still has a long way to go.

Based on interviews with experts from the Smart City Operation Center in W City and on-site field surveys, this article summarizes six problems in the digital asset information management of smart cities. Whether these issues fit with the current status of digital asset management in smart city operation centers, this article, based on interviews, invited 5 experts in related research fields to evaluate the problems identified in the digital asset information management of smart cities. The scoring is from 1 to 10 points, with 1 being the most non-conforming and 10 being the most conforming. Table 3 shows the problems existing in the information management of digital assets in smart cities.

4.2. Countermeasures for Smart City Digital Asset Information Management. This research takes the smart city of W city as an example, and proposes solutions to the existing problems in the digital asset information management of smart cities. The effect of these countermeasures needs to be

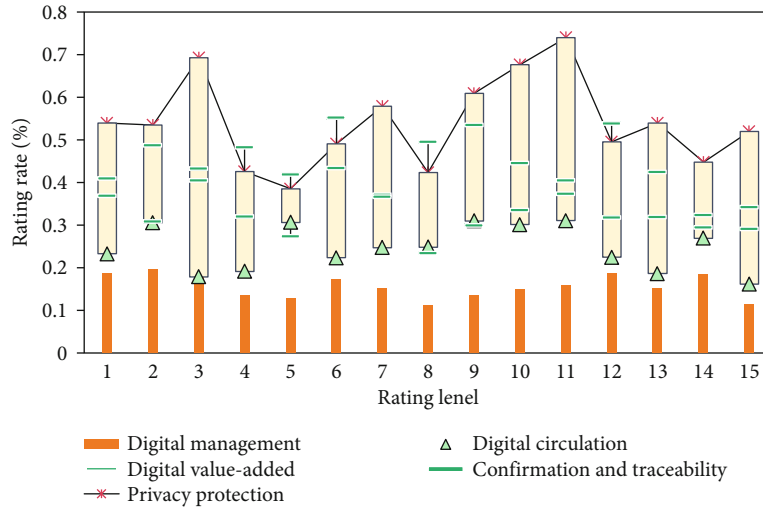


FIGURE 3: Evaluation result of expert scoring method.

analyzed. Since these are specific countermeasures and suggestions, there is no practical case to support them, so consider using the expert scoring method to analyze their effects. In the research, ten experts in related fields were invited to rate the effects of the proposed solutions (scores range from 1 to 5, with 1 being insignificant, 3 being effective, and 5 being significant).

Smart city refers to the combination of urban composition systems and services with various information technologies or innovative ideas to improve resource utilization efficiency, optimize urban management and services, and improve the life quality of citizens. Smart cities make full use of the new generation of information technology in all walks of life in the city, and build an advanced form of urban informatization on the basis of the next generation of innovation in the knowledge society (Innovation 2.0). It enables the comprehensive integration of information technology, industrialization, and urbanisation, thus alleviating “big city illnesses” and promoting urban management performance. The countermeasure of combing digital assets for digital asset audits has a high score, indicating that most experts believe that based on the current status of the development of smart city digital asset information management, it is necessary to conduct digital asset audits to further sort out the smart city operation center. Digital assets. On the whole, basically all the countermeasures can promote the management of smart city digital assets. Among them is a countermeasure, that is, whether the establishment of personnel recognition system can help break the data barrier, and the score is only 2.6. This shows that most experts believe that this countermeasure may be effective, and have no confidence in its effect. In subsequent exchanges, some experts said that data barriers exist widely on the one hand, and on the other hand, there are many influencing factors. From the perspective of the recognition system alone, the data barrier phenomenon of smart cities can only play a certain mitigation effect, and the impact capacity is limited. More effective measures should be taken, such as administrative

```
>> clear;
A=[1 1/2 1/2 1/5 1/2 1/5
  2 1 2 1/2 1/3 1/2
  2 1/2 1 1/3 1/2 1/3
  5 2 3 1 2 2
  2 3 2 1/2 1 1/3
  5 2 3 1/2 3 1];
[m,n]=size(A)
```

FIGURE 4: The constructed judgment matrix.

intervention and multi-party consultation. The evaluation result of the expert scoring method is shown in Figure 3.

Invite 10 experts in the direction of smart city and digital asset information management, and the experts will assign the above six indicators pair by pair according to the nine-level scale method, and then find the arithmetic average value to establish a judgment matrix. Bring the constructed judgment matrix into MATLAB software to check the consistency of the matrix, and calculate the weight of each index for the matrix that passes the check. The constructed judgment matrix is shown in Figure 4.

From this, the analysis of the weight of the development of digital asset information management in smart cities is shown in Figure 5. The running results of the software are as follows, the consistency index $CR=0.049$, $CI=0.0608$, and the maximum characteristic value $B=6.3040$. $CR<0.1$, it can be considered that the judgment matrix passes the consistency test and is suitable for further analysis. The weight vector Q of each indicator is: $Q=(0.0564, 0.1177, 0.0842, 0.3085, 0.1655, 0.2678)$ From this, the final weight result can be obtained. The weight value represents the weight division of experts on the further development of smart city digital asset information management. Through the above analysis, it can be found that the focus of the future development of digital asset information management in smart cities should be the data circulation link, followed by digital asset value evaluation and data appreciation. This shows that the existing smart city digital asset information management is relatively lacking in the data circulation link.

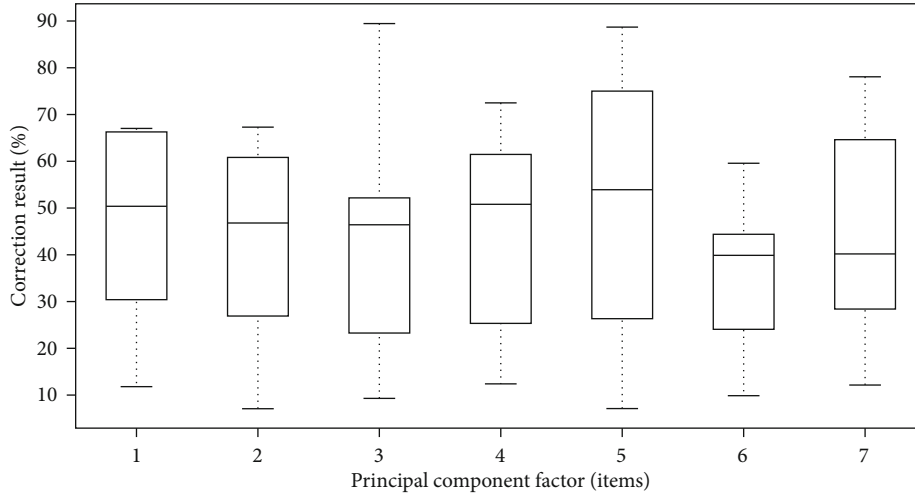


FIGURE 5: Development weights of digital asset information management in smart cities.

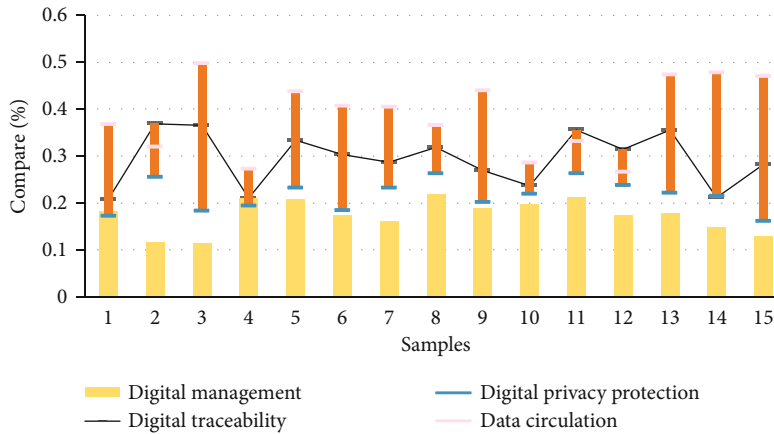


FIGURE 6: Comparison results of various index systems.

Data circulation is the basis for the realization of the value of digital assets in smart cities. At this stage, the management of many smart city operation centers does not pay attention to data circulation. Regarding how to do a good job in data circulation, this article believes that the first is to improve the convenience of data use. In the data circulation link, it is necessary to provide data products that can meet their needs from the perspective of potential data users. On the one hand, it is necessary to ensure the quality of data products, and on the other hand, to make the use of data products as convenient and fast as possible. The second is to strengthen the construction of its own big data platform and actively participate in external big data platforms.

10 appraisal experts were given scores for the judgment value of 9 relative importance in 2 levels constructed by the smart city model for the first round of scoring. The descriptive statistical analysis of the survey data eliminated 39 minor factors. Through the further investigation of the remaining 25 factors, some changes in the status of the index system were made. Simultaneously, the coefficient level has been revised by the original 22-factors in 8-elements and the coefficient level has been revised by the original 72-coefficient in 28-elements As can be observed, there are eight

factors with higher than 2.0 eigenvalues, i.e., 20.828, 2.088, 2.080, 2.688, 2.898, 2.828, 2.288, and 2.282. Those eight common factors exported can explain the 92.8% variation. The comparison result of each index system is shown in Figure 6.

5. Conclusion

This paper mainly studies the construction of digital asset in smart city, where informatization process is supported by relevant tools to meet the market demand. With respect to data quality and safety, data management and control standards must kick in to guarantee high standard and high-quality data collection. Not only should the people responsible for digital asset management be acquainted with the process of smart city informatization, but they should also be able to recognize hazards associated with smart city digital assets. Strict accountability mechanisms may protect data providers' privacy, remove certain data providers' concerns, and contribute to the lowering of data barriers. This work demonstrates the value of the proposed smart city digital asset construction scheme by comprehensive numerical results.

Data Availability

The author keeps the analysis and simulation datasets, but the datasets are not public.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by Natural Science Foundation of Shanxi Province, China (Grant No. 2014JM9371); Research on reachability of home care model in old urban residential areas in Shaanxi Province, From the perspective of digital Asset, bureau level project (21JZ034).

References

- [1] C. Zhu, L. Shu, V. C. M. Leung, S. Guo, Y. Zhang, and L. T. Yang, "Secure multimedia big data in trust-assisted sensor-cloud for Smart City," *IEEE Communications Magazine*, vol. 55, no. 12, pp. 24–30, 2017.
- [2] I. A. T. Hashem, V. Chang, N. B. Anuar et al., "The role of big data in smart city," *International Journal of Information Management*, vol. 36, no. 5, pp. 748–758, 2016.
- [3] 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.
- [4] H. Menouar, I. Guvenc, K. Akkaya, A. S. Uluagac, A. Kadri, and A. Tuncer, "UAV-enabled intelligent transportation Systems for the Smart City: applications and challenges," *IEEE Communications Magazine*, vol. 55, no. 3, pp. 22–28, 2017.
- [5] J. M. Schleicher, M. Vogler, S. Dustdar, and C. Inzinger, "Enabling a Smart City application ecosystem: requirements and architectural aspects," *IEEE Internet Computing*, vol. 20, no. 2, pp. 58–65, 2016.
- [6] F. Paganelli, S. Turchi, and D. Giuli, "A web of things framework for RESTful applications and its experimentation in a Smart City," *IEEE Systems Journal*, vol. 10, no. 4, pp. 1412–1423, 2016.
- [7] M. Gasco-Hernandez, "Building a Smart city," *Communications of the ACM*, vol. 61, no. 4, pp. 50–57, 2018.
- [8] Z. Lv, T. Yin, X. Zhang, H. Song, and G. Chen, "Virtual reality Smart City based on WebVRGIS," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1015–1024, 2016.
- [9] K. Zhang, J. Ni, K. Yang, X. Liang, J. Ren, and X. S. Shen, "Security and privacy in Smart City applications: challenges and solutions," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 122–129, 2017.
- [10] L. G. Anthopoulos and C. G. Reddick, "Understanding electronic government research and smart city: a framework and empirical evidence," *Information Polity*, vol. 21, no. 1, pp. 99–117, 2016.
- [11] M. Pouryazdan and B. Kantarci, "The smart citizen factor in trustworthy Smart City Crowdsensing," *IT Professional*, vol. 18, no. 4, pp. 26–33, 2016.
- [12] C. C. Snow, D. D. Hakonsson, and B. Obel, "A Smart City is a collaborative Community," *California Management Review*, vol. 59, no. 1, pp. 92–108, 2016.
- [13] P. Van Waart, I. Mulder, and C. De Bont, "A participatory approach for envisioning a Smart City," *Social Science Computer Review*, vol. 34, no. 6, pp. 708–723, 2016.
- [14] L. Anthopoulos, M. Janssen, and V. Weerakkody, "A unified Smart City model (USCM) for Smart City conceptualization and benchmarking," *International Journal of Electronic Government Research*, vol. 12, no. 2, pp. 77–93, 2016.
- [15] M. Guo, Y. Liu, H. Yu, B. Hu, and Z. Sang, "An overview of smart city in China," *Communications China*, vol. 13, no. 5, pp. 203–211, 2016.
- [16] D. Jin, C. Hannon, Z. Li et al., "Smart street lighting system: A platform for innovative smart city applications and a new frontier for cyber-security," *The Electricity Journal*, vol. 29, no. 10, pp. 28–35, 2016.
- [17] M. Hasan and B. Starly, "Decentralized cloud manufacturing-as-a-service (CMaaS) platform architecture with configurable digital assets," *Journal of Manufacturing Systems*, vol. 56, no. 2, pp. 157–174, 2020.
- [18] A. D. Popescu, "Transforming capital markets by means of financial digital assets," *Emulations - Revue de sciences sociales*, vol. 68, no. 2020, pp. 109–119, 2020.
- [19] S. V. Erp, "Access Management of Digital Assets," *European Property Law Journal*, vol. 8, no. 3, pp. 227–230, 2020.
- [20] Z. Li, J. Wang, and K. Li, "Digital assets Price forecast based on POW mining mechanism," *Open Journal of Social Sciences*, vol. 7, no. 2, pp. 185–198, 2019.
- [21] N. Gouru and N. L. Vadlamani, "DistProv-data provenance in distributed cloud for secure transfer of digital assets with Ethereum Blockchain using ZKP," *International Journal of Open Source Software and Processes*, vol. 10, no. 3, pp. 1–18, 2019.
- [22] K. Arrington and K. Murray, "Architecture, Design & Engineering – archiving digital assets: past, present and future," *Archiving Conference*, vol. 2018, no. 1, pp. 139–142, 2018.
- [23] S. Begum and M. Ghouse, "Access control and revocation for digital assets on cloud with consideration for sharing," *International Journal of Engineering Research & Applications*, vol. 7, no. 5, pp. 6–9, 2017.
- [24] E. Radding and K. Facedyn, "Creating a trove of digital assets at facing history and ourselves," *Journal of Digital Media Management*, vol. 5, no. 4, pp. 336–351, 2017.
- [25] N. Finzer, "Managing digital (image) assets at Northwestern University," *Journal of digital media management*, vol. 4, no. 2, pp. 116–122, 2016.

Research Article

Evolution of Regional Economic Spatial Structure Based on IoT and GIS Service

Lei Jiang 

Henan Finance University, Zhengzhou, 451464, Henan, China

Correspondence should be addressed to Lei Jiang; jianglei@hafu.edu.cn

Received 25 April 2021; Revised 1 August 2021; Accepted 25 August 2021; Published 27 September 2021

Academic Editor: Bo Rong

Copyright © 2021 Lei 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.

Unbalanced regional development is an inevitable trend in the development of all countries in the world. The rapid development of the Internet of Things (IoT) technology has created tools for the study of regional development issues. IoT has many advantages and thus owns a very wide range of applications. This paper makes use of geographic information system (GIS) technology, which can be viewed as one of the IoT sensing information. Changes in spatial regional economic differences and space and the evolution of the structure are particularly examined by processing spatial information such as maps, analyzing phenomena and events that exist on the earth, and exploiting Kriging and inverse distance weighting (IDW). The numerical results in this paper justify that the introduction of GIS technology to the study of economic diversity can upgrade regional economic research from a traditional qualitative and statistical level to a quantitative and spatial visualization level.

1. Introduction

1.1. Background and Significance. With the strong promotion of information technology, the Internet of Things (IoT) has gradually been applied in many aspects all over the world. The scale of the IoT industry has also continued to expand, becoming a new strategic industry. The IoT is not only a representative of a new generation of information technology but also an important development direction of a new generation of information technology. The IoT has many advantages, has a very strong permeability, and has a very wide range of applications. There is a close relationship between the IoT and the regional economy, which can promote the development of the regional economy, promote the transformation of the regional economy, and accelerate the growth rate of the regional economy. Therefore, research on regional economic differences based on the IoT is of great significance. Regional economic difference refers to the imbalance of the overall level of economic development between regions in a certain period of time [1]. Due to the imbalance of resources and development levels in each city (province, state), economic development cannot reach the same level in the same period, so there are regional economic

differences [2]. Regional economic disputes are a global law in the process of regional development and a key issue in regional economic research. The study of regional economic spatial differences and their causes is helpful to understand the status quo of regional economic development, promote the economic development of underdeveloped areas, and consolidate the economic achievements of developed areas [3].

The intervention of Electronic System Design Automatic-Geographical Information System (ESDA-GIS) analysis method covers the shortcomings of traditional economic difference analysis, so it is possible to study the economic spatial relationship between regional units [4]. In addition, when selecting economic indicators for European Food Safety Authority analysis, this article did not use per capita GDP indicators, but selected 13 indicators representing different levels of economic disparity for the analysis of key factors, and narrowed the scope of the 13 economic disparity indicators. Through the weighted calculation obtained by processing, the final total evaluation value of the principal component analysis is used as the county-level unit economic evaluation score of the year, as a measure of ESDA-GIS analysis, and used to analyze the subsequent county-level economic spatial correlation [5, 6].

Due to the importance of national economic research, more and more research teams have devoted themselves to the research of the national economy and have achieved very good results. For example, Farah bakhsh conducted a detailed study on the evolution of the regional economic spatial structure through the GIS method and from this infers the future economic development trend, but because it did not integrate the global environment, the conclusion is inaccurate [7]; Chhetri economy can be used to analyze specific economic conditions, but it only represents the direction of big data, and it is still not applicable to some scenarios [8]. The accuracy of economic research is very difficult [9, 10].

In a bid to improve the accuracy of the regional economic spatial structure, this paper makes use of GIS-based IoT service to study the geographical environment of the region. In particular, we employ interpolation and local fitting approaches to ensure the accuracy of local economic data. As a result, a detailed division of the regional economic spatial structure was finally drawn out through controlled experiments.

The rest of the paper is organized as follows. Section 2 presents the data processing and analyzing approaches used in this paper, including interpolation-based overall/local fitting, correlation analysis, and spacial clustering. Section 3 demonstrates the general numerical results on the evolution of the regional economic spatial structure, whereas Section 4 focuses on those specifically related to the GIS service. Finally, Section 4 concludes the paper.

2. GIS-Based Regional Economic Spatial Structure

2.1. Interpolation Method Based on Overall Fitting Technology. The entire placement technique, the placement model, is determined by all the characteristic observations of all sampling points in the target area [11]. The characteristic of this interpolation technique is that it cannot provide the local characteristics of the interpolation area, so the model is mainly used for large-scale changes [12, 13]. What we usually call the surface stress analysis method is to approximate the general trend of the sampled data by selecting a binary function [14]. The general form of the binary function is

$$F(X, Y) = \sum_{R+s=0}^{R+s=P} B_{Rs} X^R Y^s. \quad (1)$$

$F(X, Y)$ is the actual observed data value, B_{Rs} is the fitted value of the trend surface, and P is the trend surface; when $p = 0$, it is the horizontal plane:

$$F(X, Y) = B_0. \quad (2)$$

When (X, Y) changes in space, when $p = 1$, it is an inclined plane, and $B_0, B_{1X}, \dots, B_{XY}$ are the coefficients of the polynomial:

$$F(X, Y) = B_0 + B_{1X} + B_{2Y}. \quad (3)$$

When $p = 2$, it is a quadric surface:

$$F(X, Y) = B_0 + B_{1X} + B_{2Y} + B_{3X}^2 + B_{4XY} + B_5 Y^2. \quad (4)$$

Independent variable X, Y , dependent variable Z . The binary function must satisfy the least square sum of the difference between the observed value and the fitted value:

$$F(X, Y) = \arg \min \sum_{I=1}^{I=N} (Z(X_I - Y_I) - F(X_I - Y_I))^2. \quad (5)$$

Multiple regression techniques can be used to determine the aforementioned types of coefficients [15]. The use of a binary function to process surface interference voltage has the following characteristic: when $p > 3$, custom surfaces usually produce abnormally large or small values.

The placement residual is an independent error of the normal distribution and has a certain correlation. Before being used for local interpolation, the macro abnormal sampling value must be processed in advance. Total spatial interpolation was performed according to the empirical formula of one or more spatial parameters. This empirical equation is called the transformation function [16, 17], and it is also a common method of total interpolation technique. Since this article does not use placement technology, it will not be described in detail here.

2.2. Interpolation Method Based on Local Fitting Technology.

The real surface of continuous space is difficult to express with mathematical polynomials. Therefore, the local placement technique is usually used to match the value interpolation through local sampling points [18]. The positioning method only uses neighboring values to estimate the value of unknown points. Generally, these are the following steps: (1) specify the adjacent search area or range, (2) search for points located in the adjacent area, (3) choose mathematical functions that express the spatial variation of these finite points, and (4) assign values to data points belonging to ordinary grid cells. Reset the operation of this step until all points on the grid are mapped [19].

Spline interpolation is a process of obtaining a set of curve functions by mathematically solving three bending moment equations through a series of smooth curves of shape values [20]. And, when all $p - 1$ degree derivatives and adjacent blocks at the limit of n degree polynomial are continuous, it is called a spline function [13]. The principle of the mobile placement method is to match the surrounding data points by defining appropriate local functions and solving the assembly function to find the interpolation of unspecified points. This method uses an unspecified point as the interference center [21]. The Kriging (lattice) interpolation method, cofounded by French geologist Georges Matheron and South African mining engineer DG Krige in 1997, is a geostatistical method and the best linear unbiased interpolation estimator (referred to as BLUE) [22, 23]. Kriging put it this way: Suppose that Z is a regionalized variable carried by a point and is 2nd order stationary (or intrinsic), $z(X_I) (I = 1, 2, \dots, N)$, point bearer $X_I (I = 1, 2, \dots, N)$. Now,

we need to estimate the regionalization variables of the X_0 point bearing location, and the estimated amount used is

$$z'(X_I) = \sum_{I=1}^N \lambda_I z(X_I). \quad (6)$$

Choose λ_I to make the estimate of $z'(x)$ unbiased, and make the variance smaller than the variance of any linear combination of observations. That is, it satisfies the best: the deviation of the difference between the interpolation value and the true value is the smallest [24]; that is, the deviation of the difference between the interpolation value and the true value is the smallest, namely,

$$\text{Var} \left[z'(X_0) - z(X_0) \right] = \min. \quad (7)$$

Linear. The interpolation value is a linear combination of observations, namely,

$$Z'(X_0) = \sum_{I=1}^N \lambda_I z(X_I). \quad (8)$$

Unbiased Estimation. The expected value of the difference between the interpolated value and the observed value is zero, namely,

$$e \left[z'(X_0) - z(X_0) \right] = 0. \quad (9)$$

In the MAPGIS software, we have provided us with three variable function models, namely, the power exponential model, the linear model, and the spherical (MATLON) model [25].

(1) Linear model:

$$\begin{cases} \gamma(h) = c_0 + c * \left(\frac{H}{A} \right), & 0 < H < A, \\ \gamma(H) = c_0 + c, & H > a \end{cases} \quad (10)$$

(2) c is the polynomial coefficient, power exponent model:

$$\gamma(H) = c_0 + c \left[1 - \exp \left(-\frac{H}{A} \right) \right], \quad H > 0 \quad (11)$$

(3) Spherical model:

$$\begin{cases} \gamma(H) = c_0 + c \left[1.5 * \left(\frac{H}{A} \right) - 0.5 * \left(\frac{H}{A} \right)^3 \right] & 0 < H \leq A, \\ \gamma(H) = c_0 + c, & H > A \end{cases} \quad (12)$$

The range of influence or fragmentation effect is represented by the variable a , and the critical change value is rep-

resented by the variable C . Meteorologists and geologists proposed the inverse distance weighting method, which was eventually called the Shepard method [26]. The concept is to place n points, the plane coordinate (X_j, Y_j) is the vertical height, $N = 1, 2, \dots, I$, and the reciprocal distance weighted interpolation function is

$$F(X, Y) = \begin{cases} \frac{\sum_{j=1}^N (Z_j / D_j^p)}{\sum_{j=1}^N (1 / D_j^p)} & d = (X, Y) \neq (X_j, Y_j), I = 1, 2, \dots, N, \\ Z_I(X, Y) = (X_j, Y_j), & I = 1, 2, \dots, N. \end{cases} \quad (13)$$

Among them, $D_j = \sqrt{(X - X_j)^2 + (Y - Y_j)^2}$ is the horizontal distance from point (X, Y) to point (X_j, Y_j) , $J = 1, 2, \dots, N$. P is a constant greater than 0, called the weighted power exponent.

The advantage of this method lies in the fact that the formula is relatively simple, especially suitable for scattered nodes, not a problem of grid points. Its disadvantage is that it can only get the maximum and minimum values of the function at the node, as interpolation takes the weighted average of the values at each node.

2.3. Correlation Analysis Method. There are some connections between many phenomena in nature. The relationship between the above two or more random variables is determined based on mathematical statistics and is called approximate relationship or correlation. The analysis and determination of this relationship are called correlation analysis [27].

The main task of correlation analysis is to study the closeness of the relationship between variables and to draw conclusions about whether the population is relevant based on the data sample. If we can get any information about another variable from a known variable, then these two related variables are called "independent variables." The correlation between two variables may be due to various complicated reasons, or one variable affects another variable, or there is an interaction between two variables, or there is no direct relationship between two variables; all variables are also affected by the third variable. In short, the relationship between the two involves certainty and random fluctuations. In the correlation model, both variables are random variables [28]. According to the closeness of the relationship between variables, correlation types can be divided into three types. That is, complete correlation, zero correlation, and statistical correlation.

A complete correlation (functional relationship) is between two variables x and y . If there is a correspondingly defined value y for any given value x , then the relationship between the two variables is complete correlation. Zero correlation (no relationship) is when there is no relationship between two variables or the change of one phenomenon (variable) does not affect the change of another phenomenon (variable). This relationship is called zero correlation or no relationship. If the relationship between two variables

TABLE 1: Descriptive results of spatial data.

Description item	Quantity	Very bad	Minimum	Max	Avg	Standard deviation	Variance
Longitude	181	9.845	98.098	107.943	103.962	1.926	3.709
Latitude	181	7.105	26.475	33.58	30.135	1.518	2.305
GDP per capita in 2015	181	6680.936	882.8	17563.736	3243.88	2551.923	512308.582
GDP per capita in 2016	181	2620.411	1543.539	24163.95	4986.854	4192.726	7578952.143
GDP per capita in 2018	181	3804	199	7003	10154.331	7160.394	1271243.012
GDP per capita in 2019	181	1738	617	5355	12376.232	8337.462	9513280.09

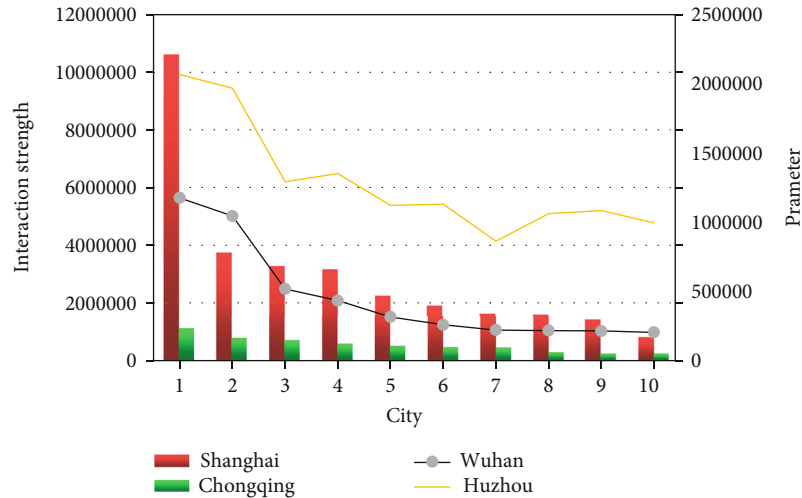


FIGURE 1: Interaction strength values of central cities in the three major regions.

is between complete correlation and zero correlation, it is called correlation or statistical correlation. When only studying the correlation between two variables, it is called simple correlation; when studying the correlation between three or more variables, it is called multiple correlation. In mathematical statistics, the parameters that determine the degree of a close correlation between variables mainly include covariance and correlation coefficient.

2.4. Spatial Clustering Method. In actual work, we often encounter the problem of sampling samples (or marking), and classification research is the basic method of scientific research. In statistics, cluster analysis is usually used to classify categories. The principle is to first treat a certain number of samples or indicators as one category and then divide the two categories by the highest affinity according to the degree of relevance of the sample (or indicator) and then consider the degree of the combined category and the other categories before the combination intimacy between. Repeat this process until all samples (or tags) are combined into one category. The spatial grouping method is different from the traditional statistical grouping analysis. First, spatial grouping is mainly based on the spatial location of geographic phenomena and reference-related feature information for grouping analysis; second, the purpose of spatial grouping is to analyze the spatial aggregation of spatial objects and their division into different subcomponents, groups (clas-

ses), and different subgroups (classes) occupying different spatial areas. The formation of subgroups is a product of the geographic environment. Based on this, certain geographic mechanisms can be revealed, and they can also be used as the basis for other analyses. Third, spatial cluster analysis is different from traditional cluster analysis. It is based on the spatial correlation of geographic variables, while spatial cluster analysis can be based on spatial autocorrelation [29].

The spatial grouping analysis method used in this document belongs to spatial statistics, and the starting point of spatial statistics is to consider that the specific geographic phenomenon or specific feature value in the peripheral unit is related to the same phenomenon or feature value in the area and adjacent area units. Spatial location produces two types of spatial effects: spatial dependence and spatial heterogeneity. The former is usually also called spatial autocorrelation or spatial correlation. Similar values in variables tend to appear in nearby locations, leading to spatial grouping. For example, some high crime areas in cities are usually surrounded by other high crime areas.

3. Numerical Results on the Evolution of Regional Economic Spatial Structure

3.1. Research Experiments on Spatial Data. In order to deepen the understanding of the internal relationship of

TABLE 2: Construction of economic difference evaluation index system.

	Selected indicator	Comment
1	Regional GDP	Sum of added value of all industries
2	GDP growth rate per capita	Important indicators of the national economy, reflecting the level of economic benefits
3	GDP growth rate	Reflect the level of economic growth
4	Industrial value added above designated size	Added value in the production process of large industrial enterprises
5	Industrial added value as a proportion of GDP	The importance of large industrial enterprises in the economy
6	The proportion of tertiary industry output value in GDP	Service-led economic transformation degree
7	The proportion of the added value of the primary industry in the area of commonly used cultivated land	Industrial productivity in the primary industry
8	Total retail sales of consumer goods per capita	Realization of the purchasing power of social goods
9	General budget revenue of local finance	Reflect the level of economic scale
10	Per capita net income of rural residents	Income of rural residents
11	Per capita net income growth rate	Income of urban residents

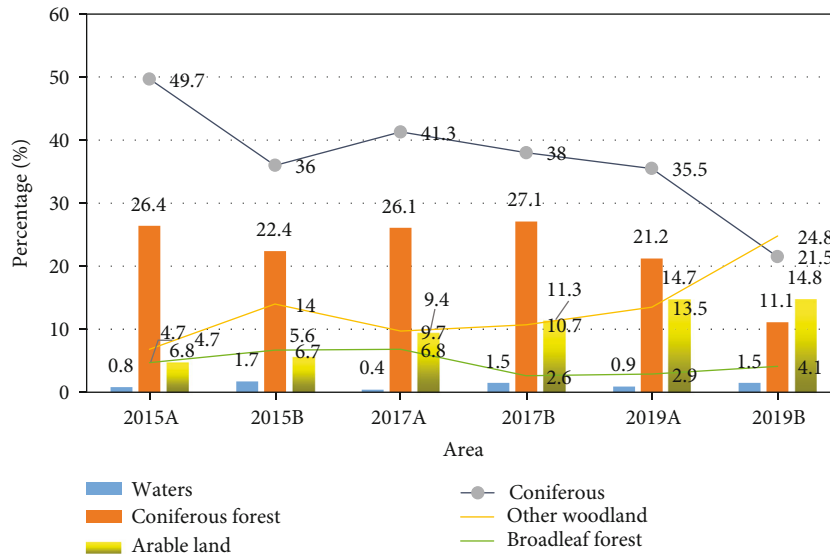


FIGURE 2: Percentage of (A, B) land use landscape type.

the data, for better and in-depth analysis of the data, recent research and analysis of spatial data are often conducted to obtain the value of spatial attributes and the spatial distribution of data. For data association, it is very useful to understand the particularity of the data, which lays the foundation for the final GIS spatial analysis research. Here, I use the corresponding SPSS statistical tools to explore and analyze the data. SPSS is the English abbreviation for Statistics Product and Service Solutions, which is a statistical software package for social sciences. It is one of the most famous statistical analysis software in the world. First, we investigate and analyze the distribution and eigenvalues of spatial data. Here, we use the analysis and description function in SPSS software. The description of the data is shown in Table 1 and Figure 1.

In addition to the variance and standard deviation in the table, other indicators are relatively easy to understand.

Among them, variance and standard deviation are descriptions of deviation trends. Deviation trend refers to the characteristics of the data set, which deviates from the central value of the distribution, reflecting the degree to which the value of each variable deviates from its central value. Through the comparative analysis of variance and standard deviation, if the variance and standard deviation of a given data set are the smallest, it means that the difference of the data set is the smallest, so the data set is more representative than some study prospects.

3.2. Experiments on the Spatial Structure and Characteristics of the Urban System. The urban system refers to a group of interconnected and evenly spaced towns in a relatively integrated region or country. Its characteristics are different types and clear division of labor. Due to regional differences and different natural environments, each city

TABLE 3: List of interaction strength values of three major regional central cities.

Serial number	Shanghai		Chongqing		Wuhan	
1	Suzhou	10609688.7	Chengdu	1101360	Ezhou	1175061.88
2	Wuxi	3749384.92	Luzhou	770543.54	Xiaogan	1044256.58
3	Hangzhou	3290657.92	Guangan	695230.8	Huanggang	517237.24
4	Nantong	3174218.88	Nanchong	565544.55	Huangshi	432841.85
5	Ningbo	2265610.52	Neijiang	487017.88	Xianning	315264.27
6	Changzhou	1919727.61	Zigong	446243.63	Changsha	258717.88
7	Nanjing	1649481.85	Suining	431156.91	Nanchang	220375.88
8	Jiaxing	1611413.01	Yibing	268598.96	Shanghai	217113.32
9	Shaoxing	1395929.23	Guiyang	228169.92	Nanjing	213876.45
10	Huzhou	785746.26	Ziyang	222364.65	Yueyang	203564.99

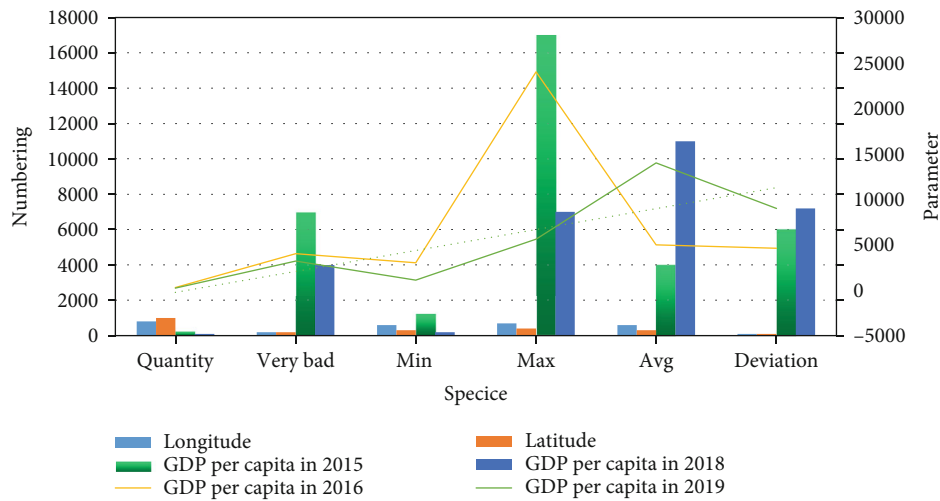


FIGURE 3: Descriptive results of spatial data.

will develop in a direction conducive to its development. As time goes by, urban agglomerations will form a clear division of labor and interconnections, for example, traffic cities, tourist cities, historical and cultural cities, coal cities, and complete cities. It is variable. The urban system is not static after its formation. The scale, structure, and form of the urban system will change over time and change in government planning. You must have integrity. The bourgeois system is not a closed social and economic system. It refers to a unified whole composed of a series of cities of different scales, different functions, and interconnected cities, with the designated central city as the core of the defined area. The cycle and exchange of energy, matter, and information continue with each other. The bourgeoisie and the outside world continue to exchange and cooperate in the fields of politics, economy, culture, science and technology, trade, etc., in order to strengthen the external relations of the bourgeois system and its own rapid and healthy growth.

The spatial structure of the urban system refers to the spatial interaction of towns in a region, which merges the spatially separated towns into an organic whole with a spe-

cific structure and function. Studies have shown that the spatial distribution of urban systems has obvious scale-free characteristics and has a random fractal structure within a certain range. There are three basic fractal dimensions to describe the spatial structure characteristics of the peripheral urban system: one is the fractal set dimension, which starts from the point density and describes the same characteristics of the urban peripheral spatial distribution; the other is the dimension fractal correlation from multiple starting with point density; it describes the relative distribution of system components; the third is the fractal network dimension, which starts directly from the data distribution and describes the spatial structure of the system.

With the rise of transportation and information technology, the spatial structure of today's global system is based on the connections of channels, nodes, levels, flow, terrain, and networks. Nodes and channels are the material basis of the space structure. Domain and network are the functional elements that form the spatial characteristics. It can be considered that the spatial structure of the urban system is composed of different levels of cities, spatial flows, passages, regions, and networks.

TABLE 4: Percentage of (A, B) land use landscape type.

Land landscape Type	2015		2017		2019	
	A	B	A	B	A	B
Waters	0.8	1.7	0.4	1.5	0.9	1.5
Coniferous forest	26.4	22.4	26.1	27.1	21.2	11.1
Coniferous and broad-leaved mixed forest	49.7	36	41.3	38	35.5	21.5
Other woodland	6.8	14	9.7	10.7	13.5	24.8
Arable land	4.7	5.6	9.4	11.3	14.7	14.8
Broadleaf forest	4.7	6.7	6.8	2.6	2.9	4.1

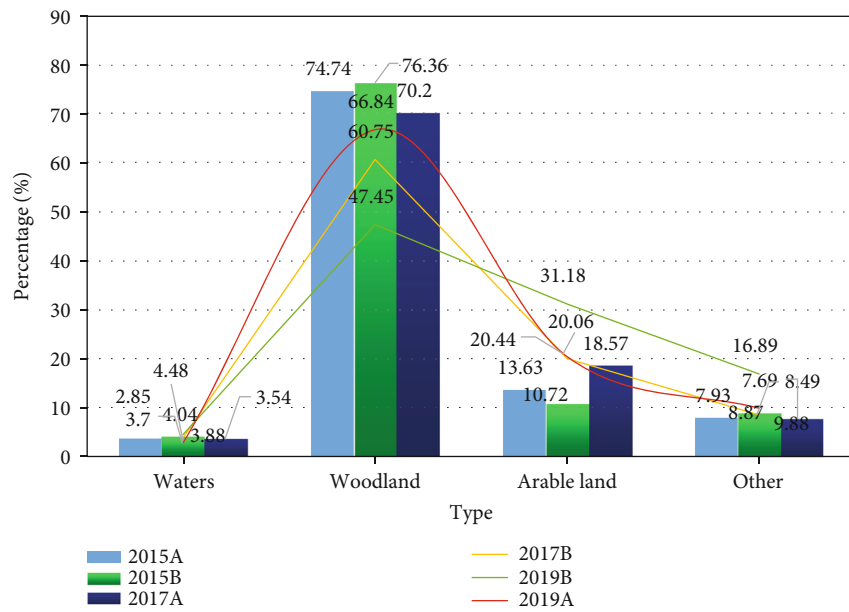


FIGURE 4: Percentage of land use landscape types in A and B.

TABLE 5: Percentage of land use landscape types in A and B.

Land landscape type	2015		2017		2019	
	A	B	A	B	A	B
(1) Waters	3.7	4.04	3.54	3.88	2.85	4.48
(2) Woodland	74.74	76.36	70.2	60.75	66.84	47.45
(3) Arable land	13.63	10.72	18.57	20.06	20.44	31.18
(4) Other	7.93	8.87	7.69	8.49	9.88	16.89

4. Evolution of Regional Economic Spatial Structure Based on GIS Location Services

4.1. Interaction Between Cities in the Study Area. The law of universal gravitation is the law of gravitation that explains the interaction between objects. It is the law of mutual attraction between objects due to their mass. Since the urban system is located in a relatively integrated region or country with the central city as the core, it is composed of a series of cities with different scales, different functions, and close connections. At the same time, the urban system is complete, hierarchical, and dynamic. The internal relationship of the urban system is like every planet has satellite

orbits, and central cities also have satellite cities, county seats, and other systems, forming a complete urban system. And there is a relationship between attraction and repulsion according to the universal law of gravitation. When a city is far away from the central city, the attractiveness and aversion of the central city are relatively weak, and the space for independent development is larger, even affected by the surrounding cities; it is also close to another system and attracted by culture and its policy. When the distance is relatively close, the impact will be greater and the resources will be relatively greater, so it will be developed and promoted in collaboration with other parts of its own system.

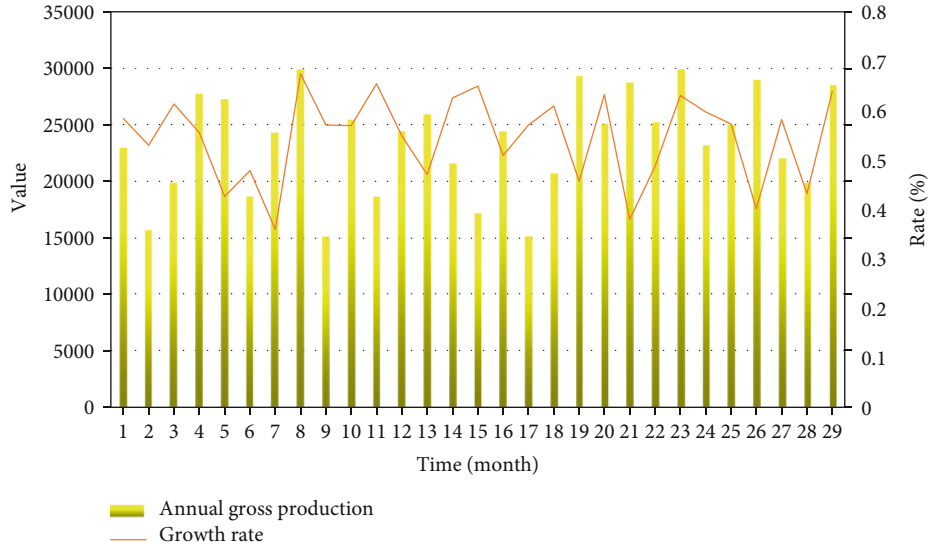


FIGURE 5: Annual production table of A.

There is a relationship between gravity and repulsion between towns. Under the influence of distance, economy, and culture, when the gravitational and repulsive forces between the two cities are balanced, the cities will develop together; otherwise, they will restrict and hinder each other and affect the city. The development of the city will lead to the unity or diversification of the city. In the urban system, due to distance, cities outside the system may be greatly affected by other systems, and the development of cities will also tend to other systems. This article introduces the types of gravity in physics to analyze the strength of the interaction between cities. Take the straight-line distance between cities as the radius, and calculate the attractiveness between the two as a feature to characterize the intensity of interaction between cities. The spatial interaction force M is used to measure the strength of the interaction between cities, the city's population size, economic development level, and other related indicators which are used as characteristic values to measure the quality of the city, and the interaction between cities is studied and calculated.

Taking into account the availability of data, the population data of 110 cities in the Yangtze River Economic Zone in 2014 and the GDP of the same year were selected, and the direct distance data between cities was used for calculation. According to the method of calculating the urban interaction force, the value of the interaction force among 110 cities in the Yangtze River Economic Zone was calculated. The results show that in the city system of the Yangtze River Economic Belt, Shanghai has the strongest interaction with other cities, the closest connection, and the highest sum of interaction possibilities, which is 36967794, which is significantly higher than other cities. He established Shanghai in the urban residential area of the Yangtze River Economic Zone. For functional and organizational key positions, except for Shanghai, the top 15 cities in terms of the intensity of interaction with

TABLE 6: Annual production table of A.

Area A	2015	2016	2017	2018	2019
Annual gross product	22990	27722	34606	37757	40154
Growth rate	8.9	11.9	8	8.2	7.6

other cities are Suzhou (31531495), Wuxi (25667714), Nanjing (22230653), Hangzhou (18759240), Changzhou (16964566), Yangzhou (14270), Shaoxing (10821511), Zhenjiang (10567366), Nantong (8900823), Chongqing (8233501), Wuhan (7601085), Ningbo (6845078), Changsha (6574185), and Taizhou (6308996). Among the 15 cities, 12 cities belong to the Yangtze River City Group, 2 cities belong to the Middle Yangtze River City Group, and 1 city belongs to the upper Yangtze River City Group. In order to facilitate the study of the overall urban spatial interaction of the Yangtze River Economic Belt, Chongqing in the upper reaches of the Yangtze River, Wuhan in the middle reaches of the Yangtze River, and Shanghai in the lower reaches of the Yangtze River were selected and their power to interact with cities in each region. Table 2 and Figure 2 show the top ten cities ranked from high to low in terms of interaction intensity with the three outer cities.

4.2. Land Use Changes on Both Sides of the Region. Based on the interpretation results of land use in different periods of the above two typical plots, the interpretation results are then counted to obtain the percentages of different land use landscape types in the two study plots.

It can be seen from Table 3 and Figure 3 that the percentage of land use types in different periods is different. Generally, coniferous forests and deciduous forests account for a larger percentage, while other types of land use account for a relatively small percentage. At the same time, the

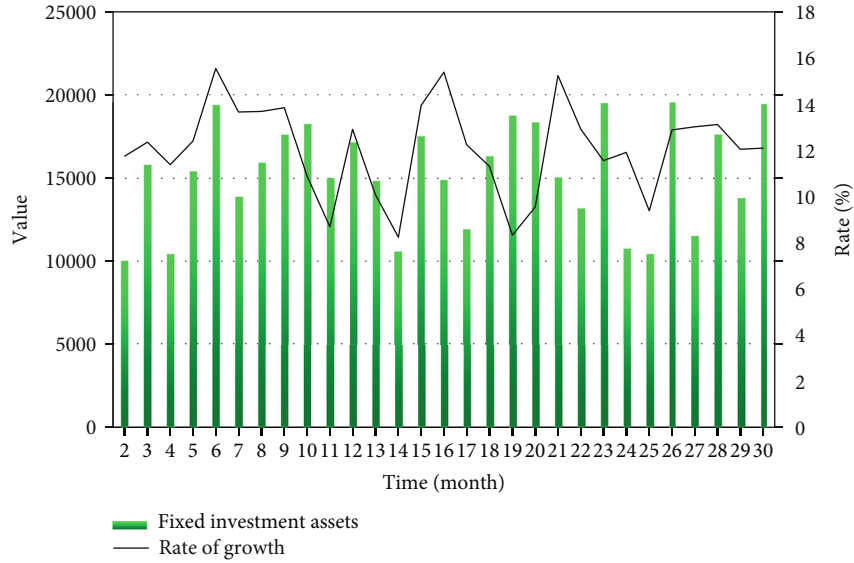


FIGURE 6: Fixed assets of the whole society in A.

analysis of the main land use landscape types on both sides of the A and B boundary shows that in all three periods, the percentage of coniferous forest on the A side is higher than that on the B side. A means decreases every year. The B end accounted for a sizable percentage, and both sides were on the rise year after year. On side B, the fraction of cultivated land is comparatively high, and both sides are rising year by year; on side A, the proportion of mixed coniferous forest is relatively high, and both sides are decreasing year by year.

Through the analysis of Table 4 and Figure 4, it can be seen that forest land is the main land use landscape type on both sides of the boundary between A and B, and forest land accounts for the largest proportion, followed by cultivated land. This is mainly due to the rehabilitation of human beings, which makes the cultivated area continue to increase. The analysis of the land use landscape types on both sides of the border between A and B in different periods shows that in all four periods, except for side B, the largest percentage of woodland in 1976, both sides accounted for a larger proportion. Both A and B sides showed a downward trend year by year; except for the larger proportion of B side in 1976, the proportion of cultivated land in other periods was higher than that of A side. The results show that the artificial arable land of side A has been restored to a large extent, and the two sides of the boundary have been increasing year by year, and the proportion of other lands on side B is relatively high, and the two sides of the boundary have a trend of increasing year by year.

4.3. *Current Economic Situation of A.* Table 5 and Figure 5 show the annual GDP of Province A from 2015 to 2019. The graph shows that the GDP of Province A has increased year by year, from 2,299 billion in 2015 to 40,154 billion yuan in 2019. In terms of the growth rate, the average annual growth rate of GDP has decreased since 2010, indi-

TABLE 7: Fixed assets and their growth rates in the whole society of A.

Area A	2015	2016	2017	2018	2019
Fixed investment assets	9906	11452	17096	20194	23555
Rate of growth	15.9	15.6	21.4	18.1	16.6

cating an overall negative trend. The GDP growth rate is increasing from a quantitative standpoint. In 2010, the economic growth rate was the fastest, reaching 11.9%, followed by the annual GDP growth rate of A. Prices rise slowly. It can be seen that in the process of rapid economic growth, Area A has encountered some bottlenecks. These bottlenecks hinder the rapid growth and slow down A region’s economy. Combined with the actual situation of A, part of the reason may be due to the environmental problems of the earth. Obstacles to economic growth, especially the land problem of A, make sustainable economic growth possible. Therefore, in the process of economic development, it is necessary to combine the actual environment of A to coordinate growth.

Table 6 and Figure 6 show the total investment in fixed assets of society A from 2015 to 2019. The picture shows that the fixed asset investment of society A has increased from 990.6 billion yuan in 2015 to 2,355.5 billion yuan in 2019, and the investment amount has almost doubled. Such a huge change is also an important reason for A’s economic growth. Although the number of fixed assets has accumulated in a large amount, it is not difficult to see from the quantity that the annual growth rate of fixed asset investment has been declining very slowly since 2017, and the downward trend is very obvious. The growth rate has begun to decline from 22.9% per year. A’s investment share will fall into a “recession” every year, as shown in Table 7 and Figure 7.

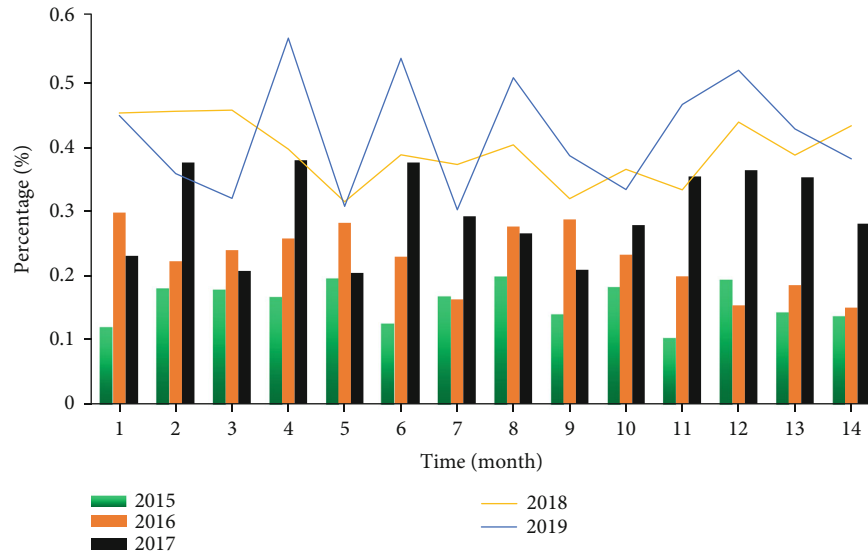


FIGURE 7: Growth rate of fixed assets and annual production growth rate in A.

5. Conclusions

This paper mainly studies the evolution of regional economic spatial structure based on GIS positioning-related IoT services. In particular, we analyze the economic growth and investigate the evolution process of the regional economic space structure. The data set used in this study is largely derived from a brief period of time on the Chinese market. Additionally, we will collect additional data over a longer period of time around the world in order to build a more universal theoretical framework in the near future.

Data Availability

All data used are given in the paper.

Conflicts of Interest

The author declares that they have no conflicts of interest.

Acknowledgments

This paper is financially supported by Henan Finance University.

References

- [1] A. Farahbakhsh and M. A. Forghani, "Sustainable location and route planning with GIS for waste sorting centers, case study: Kerman, Iran," *Waste Management & Research*, vol. 37, no. 3, pp. 287–300, 2019.
- [2] W. Li, K. Cao, and R. L. Church, "Cyberinfrastructure, GIS, and spatial optimization: opportunities and challenges," *International Journal of Geographical Information Science*, vol. 30, no. 3, pp. 427–431, 2016.
- [3] P. Chhetri, B. Kam, K. Hung Lau, B. Corbitt, and F. Cheong, "Improving service responsiveness and delivery efficiency of retail networks," *International Journal of Retail & Distribution Management*, vol. 45, no. 3, pp. 271–291, 2017.
- [4] Z. Wang, "Research of the optimization algorithm for railway spatial structure distribution based on GIS," *Shanxi Architecture*, vol. 45, no. 6, pp. 113–114, 2019.
- [5] W. Huang, B. W. Baetz, and S. Razavi, "A GIS-based integer programming approach for the location of solid waste collection depots," *Journal of Environmental Informatics*, vol. 28, no. 1, pp. 39–44, 2014.
- [6] H. Lim and M.-W. Koo, "Promoting cost efficiency and uniformity in parcel delivery centre locations and service areas: a GIS-based analysis," *International Journal of Logistics Research and Applications*, vol. 19, no. 5, pp. 369–379, 2016.
- [7] K. Tiwari, R. Goyal, and A. Sarkar, "GIS-based methodology for identification of suitable locations for rainwater harvesting structures," *Water Resources Management*, vol. 32, no. 5, pp. 1811–1825, 2018.
- [8] Z. Ding, K. Zhou, J. Kang, F. Wang, and G. Zhang, "The spatial differentiation and influencing factors of the service quality of C2C stores in Central China: a case study of five types of Taobao online stores," *Geographical Research*, vol. 35, no. 6, pp. 1074–1094, 2016.
- [9] G. Kbar, M. Abidi, S. Hammad Mian, A. al-Daraiseh, and W. Mansoor, "A university-based smart and context aware solution for people with disabilities (USCAS-PWD)," *Computers*, vol. 5, no. 3, pp. 18–23, 2016.
- [10] C. Yan-yan, W. Pan-yi, L. Jian-hui, F. Guo-chen, L. Xin, and G. Yi, "An evaluating method of public transit accessibility for urban areas based on GIS," *Procedia Engineering*, vol. 137, pp. 132–140, 2016.
- [11] L. Huang, S. Xu, L. Lu, L. Xu, and Y. Zhang, "A method of location selection for rural highway transportation service facilities based on GIS," *Journal of Agricultural Science*, vol. 12, no. 7, pp. 53–65, 2020.
- [12] P. Rathore, S. P. Sarmah, and A. Singh, "Location-allocation of bins in urban solid waste management: a case study of Bilsapur city, India," *Environment, Development and Sustainability*, vol. 22, no. 4, pp. 3309–3331, 2020.

- [13] R. L. Church and W. Li, "Estimating spatial efficiency using cyber search, GIS, and spatial optimization: a case study of fire service deployment in Los Angeles County," *International Journal of Geographical Information Science*, vol. 30, no. 3, pp. 535–553, 2016.
- [14] D. Ahmetovic, R. Manduchi, J. M. Coughlan, and S. Mascetti, "Mind your crossings: mining GIS imagery for crosswalk localization," *ACM Transactions on Accessible Computing*, vol. 9, no. 4, pp. 1–25, 2017.
- [15] L. Guo, J. Luo, M. Yuan, Y. Huang, H. Shen, and T. Li, "The influence of urban planning factors on PM_{2.5} pollution exposure and implications: a case study in China based on remote sensing, LBS, and GIS data," *Science of The Total Environment*, vol. 659, pp. 1585–1596, 2019.
- [16] Z. Ya, "Design and implementation of fire fighting digital drawing query system," *Computer Knowledge and Technology: Academic Edition*, vol. 15, no. 7, pp. 76–77, 2019.
- [17] N. Ya'acob, M. S. Ahmad Azmil, K. N. Tahar et al., "Geographical information system (GIS) map for fire and rescue application," *Jurnal Teknologi*, vol. 78, no. 5–9, 2016.
- [18] A. T. Alsharif, E. Kruger, and M. Tennant, "Identifying and prioritising areas of child dental service need: a GIS-based approach," *Community Dental Health*, vol. 33, no. 1, pp. 33–38, 2016.
- [19] A. Dejen, S. Soni, and F. Semaw, "Spatial accessibility analysis of healthcare service centers in Gamo Gofa Zone, Ethiopia through geospatial technique," *Remote Sensing Applications: Society and Environment*, vol. 13, pp. 466–473, 2019.
- [20] M. Abdulkader, "Using GIS for determining variations in health access in Jeddah City, Saudi Arabia," *ISPRS International Journal of Geo-Information*, vol. 7, no. 7, pp. 254–298, 2018.
- [21] S. Hong, "Design and development of GIS-based air quality information system for ubiquitous public access," *Journal of the Korean Society of Civil Engineers*, vol. 37, no. 1, pp. 195–201, 2017.
- [22] F. Sandrone and V. Labiouse, "A GIS based approach for analysing geological and operation conditions influence on road tunnels degradation," *Tunnelling and Underground Space Technology*, vol. 66, pp. 174–185, 2017.
- [23] M. Menon and R. Mohanraj, "Temporal and spatial assemblages of invasive birds occupying the urban landscape and its gradient in a southern city of India," *Journal of Asia Pacific Biodiversity*, vol. 9, no. 1, pp. 74–84, 2016.
- [24] R. Kato, H. Goto, and S. Yoshie, "The current installation and spacial characteristics of the "semi outdoor advertisement" -from the survey of facades in Ginza commercial district-," *Journal of Architecture and Planning (Transactions of AIJ)*, vol. 81, no. 730, pp. 2741–2751, 2016.
- [25] A. Bassanini, G. Brunello, and E. Caroli, "Not in my community: social pressure and the geography of dismissals," *Journal of Labor Economics*, vol. 35, no. 2, pp. 429–483, 2016.
- [26] E. Gildenhuis, A. Ellis, S. Carroll, and J. L. Roux, "The ecology, biogeography, history and future of two globally important weeds: *Cardiospermum halicacabum* Linn. and *C. grandiflorum* Sw," *Revista Brasileira De Botnica*, vol. 19, no. 31, pp. 45–65, 2017.
- [27] W. Zheng, Y. Jiang, R. Zhuo, J. Run, and X. Wang, "Evolution and influencing factors of the structure of economic linkage network at county level in Anhui Province," *Scientia Geographica Sinica*, vol. 36, no. 2, pp. 265–273, 2016.
- [28] S. Z. Li, "Discussion on the evolution and optimization of economic spatial structure in Linyi, Shandong Province," *Geographical Science Research*, vol. 9, no. 1, pp. 19–27, 2020.
- [29] H. Yue and L. I. Lin, "A comprehensive assessment of green development and its spatial-temporal evolution in urban agglomerations of China," *Geographical Research*, vol. 36, no. 7, pp. 1309–1322, 2017.

Research Article

BEI-TAB: Enabling Secure and Distributed Airport Baggage Tracking with Hybrid Blockchain-Edge System

Pengbo Si , Fei Wang , Enchang Sun, and Yuzhao Su

Faculty of Information Technology, Beijing University of Technology, 100124, China

Correspondence should be addressed to Pengbo Si; sipengbo@bjut.edu.cn

Received 7 July 2021; Accepted 27 August 2021; Published 23 September 2021

Academic Editor: Michel Kadoch

Copyright © 2021 Pengbo Si 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.

Global air transport carries about 4.3 billion pieces of baggage each year, and up to 56 percent of travellers prefer obtaining real-time baggage tracking information throughout their trip. However, the traditional baggage tracking scheme is generally based on optical scanning and centralized storage systems, which suffers from low efficiency and information leakage. In this paper, a blockchain and edge computing-based Internet of Things (IoT) system for tracking of airport baggage (BEI-TAB) is proposed. Through the combination of radio frequency identification technology (RFID) and blockchain, real-time baggage processing information is automatically stored in blockchain. In addition, we deploy Interplanetary File System (IPFS) at edge nodes with ciphertext policy attribute-based encryption (CP-ABE) to store basic baggage information. Only hash values returned by the IPFS network are kept in blockchain, enhancing the scalability of the system. Furthermore, a multichannel scheme is designed to realize the physical isolation of data and to rapidly process multiple types of data and business requirements in parallel. To the best of our knowledge, it is the first architecture that integrates RFID, IPFS, and CP-ABE with blockchain technologies to facilitate secure, decentralized, and real-time characteristics for storing and sharing data for baggage tracking. We have deployed a testbed with both software and hardware to evaluate the proposed system, considering the performances of transaction processing time and speed. In addition, based on the characteristics of consortium blockchain, we improved the practical Byzantine fault tolerance (PBFT) consensus protocol, which introduced the node credit score mechanism and cooperated with the simplified consistency protocol. Experimental results show that the credit score-based PBFT consensus (CSPBFT) can shorten transaction delay and improve the long-term running efficiency of the system.

1. Introduction

According to the latest data from the International Air Transport Association (IATA), global air transport carries about 4.3 billion pieces of baggage every year [1]. The annual increase of baggage brings new challenges to airports. The current status and problems in baggage tracking are as follows:

(1) Optical scanning code is widely used for baggage tracking which leads to the inability to collect real-time information of baggage handling causing passengers' anxiety for waiting and management's failure to grasp the handling situation of baggage [2]

(2) The lack of unitive platform that managed and shared processing information for each baggage leads to information islands [3]

(3) The current logistics records are recorded in centralized database, which more likely leads to passenger information leakage and tampering

Compared with optical scanning code, RFID possesses advantages of fast recognition speed, large data capacity, long service life, and reusability. Baggage tracking technology based on RFID is one of the most advanced technologies in international baggage management, but the data sharing pattern and security still need to be promoted [4].

Blockchain technology could be a promising technology that brings essential changes to air baggage tracking. Blockchain maintains and records transactions of events that are immutable and cannot be falsified. It provides transparent, secure, and trustworthy data in both private and public domains, which solves the problem of information leakage and tampering caused by single centralized database in baggage tracking systems. At the same time, in traditional ways, it is difficult to guarantee data privacy and build trust between participants in multidepartment cooperation. The decentralized nature of blockchain can efficiently establish a data-sharing model and ensure multidepartment encryption cooperation. In addition, the traceable chain structure of blockchain ensures that the data cannot be tampered with, which can significantly improve the trustworthiness of baggage tracking and retrieval. Compared with public blockchain, consortium blockchain only supports the access of the nodes participating in maintenance, and participating nodes need authorization before joining and maintaining blockchain. In addition, its authorized access features can reduce the degree of data leakage, thus enhancing the privacy security of data. Consortium blockchain can guarantee the security of data, but its scalability is challenged in the face of hundreds of millions of baggage [5], so we introduced IPFS in our system.

IPFS provides a point-to-point (P2P) distributed storage structure, which can easily store a large amount of passenger data. IPFS is a content-addressed block storage system with features such as secure transaction hash mapping, high throughput, and concurrent access to transactions by peers in the network [6]. Besides, we take advantage of CP-ABE to realize attribute-based access control.

Consensus algorithm affects the performance of blockchain system. The PBFT algorithm is mainly used in consortium blockchain to solve Byzantine general problem. However, the consistency protocol of PBFT requires to complete two times of node communication which complexity is $O(N^2)$, where N is the total number of nodes in the network, resulting in high communication complexity and cost.

Aiming at the above problems, we propose a blockchain and edge computing-based IoT system for tracking of airport baggage (BEI-TAB). Our contributions are mainly as follows:

- (1) The application of blockchain in baggage tracking not only reduces the degree of data leakage but also enhances the privacy and security of data through utilizing the features of nontampering of blockchain and authorized access of consortium blockchain
- (2) The multichannel design enables the airport to process multiple types of data and business requirements in parallel and rapidly, providing coarse-grained privacy protection and promotes information sharing. It realizes the physical isolation of data and further ensures the confidentiality of transmission
- (3) By combining RFID with the blockchain, the real-time baggage processing information is automatically stored in the blockchain, which effectively saves the labor cost as well as guarantees the safety of data transmission and improves the degree of informatization

- (4) The integration of IPFS and blockchain realizes the storage of encrypted basic baggage information in the IPFS network, while only the IPFS address hash is stored in blockchain, which increases the scalability of the blockchain system. At the same time, the application of CP-ABE takes advantage of attribute-based encryption and provides fine-grained privacy protection
- (5) Both software and hardware were deployed in a testbed to evaluate the performance of transaction processing time and speed for the proposed system
- (6) In order to simplify the communication process, in the absence of Byzantine nodes, CSPBFT adopted a simplified consistency protocol to reduce the communication traffic between nodes. Besides, the node credit score mechanism effectively identifies and excludes Byzantine nodes in the system, so that the algorithm can execute the simplified consistency protocol most of time, thus improving the long-term operation efficiency of the system

2. Related Works

In this section, we review the related work on baggage tracking. At present, the improvement of baggage handling system mainly combines new technologies such as machine vision and IoT to solve baggage transportation errors, and there are few methods using blockchain technology. Singh et al. [7] propose a design of baggage tracing and handling system using smart RFID tags and IoT which is based on cloud server. However, the baggage's real-time position is tracked and stored in a cloud, which centralized storage potentially leads to information leakage. Jerry et al. [8] propose a system based on RFID, ZigBee, and GSM to update the status of baggage at various points in the journey map. However, it does not focus on secure information transmission and sharing. Johnson et al. [9] design a machine vision-based airport baggage tracking system using an integral image to obtain the bag location, but the massive amount of information poses a challenge to the system. Gao and Liang [10] adopt a convolutional neural network with video input to detect the appearance transportability of baggage. However, problems such as baggage recovery and reliable cooperation between different departments cannot be resolved. Wang et al. [11] introduce a social network that combines the IPFS, Ethereum, and attribute-based encryption to realize the access control to the data by setting out the access policy, but it has two problems. Firstly, the security level of public blockchain is excellent, but it also has slow transaction process speed and low throughput [12]. Conversely, the nodes in the consortium blockchain do not need to keep accounts through the competitive consensus mechanism so that the consensus transaction speed is higher than public blockchain. In addition, its authorized access mechanism can reduce the degree of data leakage and thus enhances the security of privacy data. As noted above, our system chooses Hyperledger Fabric, which is one of the most popular consortium blockchain platforms, as our blockchain platform.

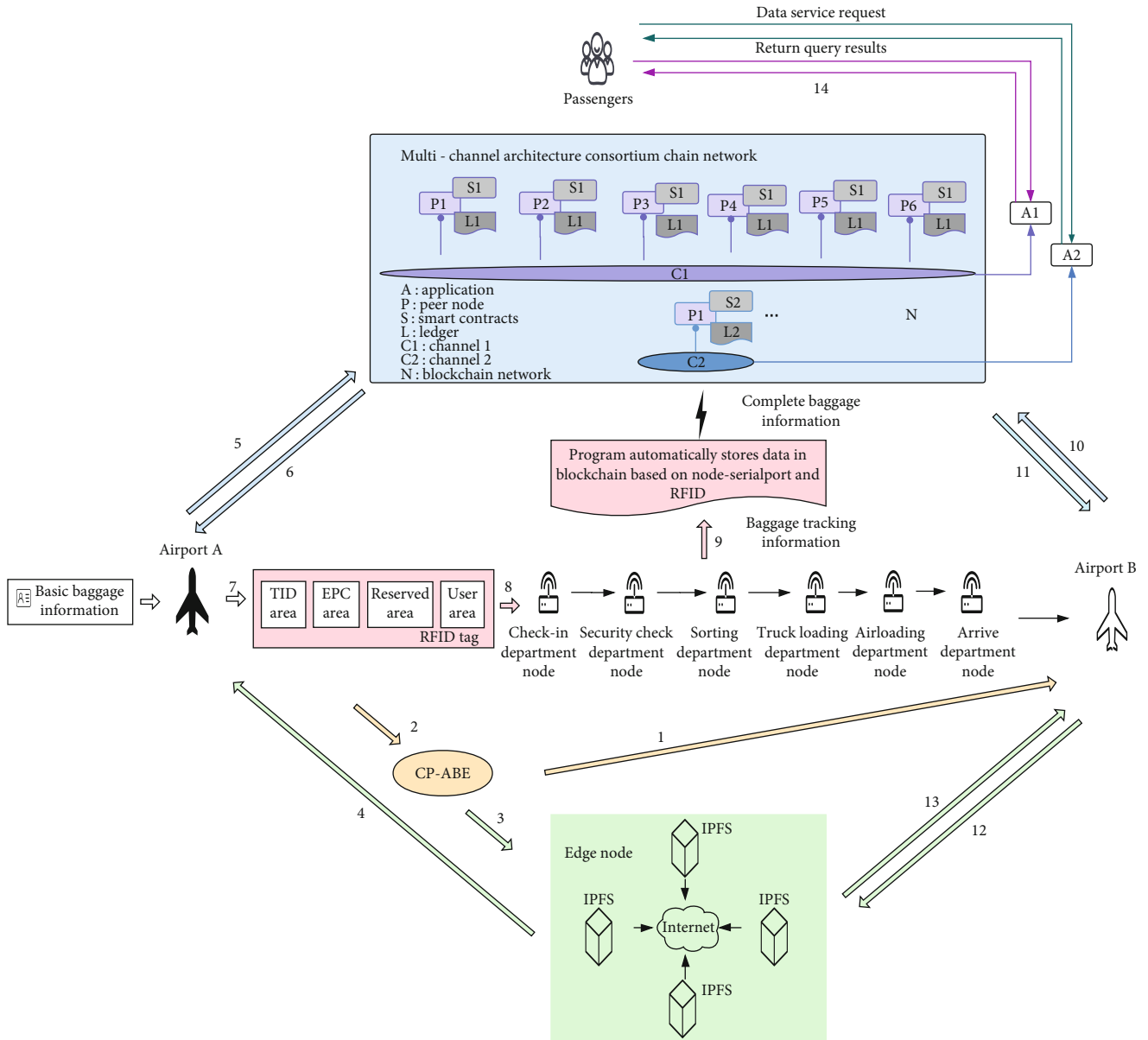


FIGURE 1: BEI-TAB architecture.

3. BEI-TAB Architecture

In this section, we illustrate the overall architecture of BEI-TAB and account for the following entities that take part in our architecture by referring to Figure 1.

Nodes. On the one hand, the nodes in consortium blockchain network refer to departments of airlines. There are six basic departments involved in baggage transport at one time: check-in, security check, sorting, trucking loading, air loading, and arrival. On the other hand, these nodes are also equipped with RFID readers. RFID readers obtain data from RFID tags by radio waves, which is a kind of automatic identification and data collection (AIDC) technology [13]. RFID system is composed of three components: RFID tags, RFID readers, and antenna. The internal storage area of RFID tags can be divided into four areas:

- (1) *Tag Identification (TID) Area.* Storing TID number, it is readable but not writable, and each TID number is unique
- (2) *User Area.* Storing user-defined data
- (3) *Electronic Product Code (EPC) Area.* Storing EPC numbers, which are unique electronic codes of objects
- (4) *Reserved Area.* Storing kill password and access password

Edge nodes. Edge nodes play the role of off-chain storage devices in our architecture. We build the IPFS network on edge nodes, which stores basic baggage information data and imposes attribute-based access control policies. IPFS is a decentralized data management system based on a P2P network model. It can connect computer devices in the same

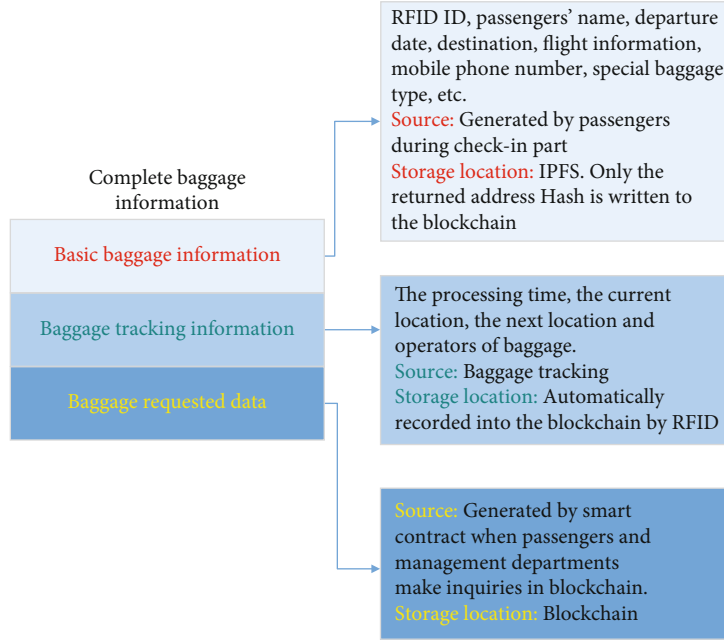


FIGURE 2: A complete piece of baggage information.

network while ensuring that files will not be tampered with [14]. IPFS stores files through content-addressed hash in a distributed hash table (DHT) which adopts version-control history to remove duplicate files. After uploading a file to IPFS, it returns the unique content-addressed hash while users only need this hash to access the resource [15].

Airport. Different airports play the role of baggage information provider or visitor. Encryption policies are implemented in CP-ABE to control business access of different airport departments to ensure decentralized and secure characteristics for basic baggage information.

Attribute-based encryption (ABE) is an access control technology. Private keys and ciphertexts in ABE are associated with the attributes of users or organizations. The resource providers only need to encrypt the message according to the attributes, no longer need to pay attention to the number or identity of the members in the group, which reduces the data encryption cost and protects the privacy of users. ABE can be divided into two categories: CP-ABE and key policy attribute-based encryption (KP-ABE). In CP-ABE, the access policy is generated by senders and bounded to the ciphertext, and private key is combined with the user's own attributes. In KP-ABE, the access policy is generated by receivers. As described above, CP-ABE is more favorable for our framework compared with KP-ABE. CP-ABE has four algorithms [16]:

$$\text{Setup}(\lambda) \longrightarrow \text{PK, MSK}. \quad (1)$$

The setup algorithm takes the security parameter λ as input and outputs the public key PK and master key MSK.

$$\text{Encrypt}(\text{PK}, p, m) \longrightarrow \text{CT}. \quad (2)$$

The encryption algorithm takes PK, message m , and an access policy p as input. It will produce ciphertext CT.

$$\text{keyGen}(\text{MSK}, S) \longrightarrow \text{SK}. \quad (3)$$

The key generation algorithm accepts the input including MSK and a set of attributes S . It creates a private key SK which is linked with attributes.

$$\text{Decrypt}(\text{PK}, \text{CT}, \text{SK}) \longrightarrow m. \quad (4)$$

The decryption algorithm receives the input including CT, p , and SK. Attributes that satisfy the policy p are able to decrypt the message. This step will decrypt the ciphertext and return message m .

Complete baggage information. A complete piece of baggage information is divided into three parts as Figure 2 showed.

Smart contract. The business logic of smart contract can be summarized as following three parts:

- (1) Combining with RFID to realize real-time automatic storage of baggage tracking information in blockchain
- (2) Providing data access and interaction interfaces for passengers and administrators, respectively. Passengers can query baggage information through RFID ID to obtain the whole process of baggage tracking, thus reducing anxiety. The administrator can carry out accurate or batch retrieval through RFID ID and flight number so that the handling status of baggage can be grasped
- (3) Generating statistics of baggage request data when passengers and management departments make real-time query requests

The smart contract interface is shown in Table 1.

TABLE 1: Smart contract interface.

Interface definition	Interface description	Function description
Query	Information query	Data sharing and retrieval
Delete	Information delete	
CreateZJxl	Check-in department information added to the blockchain	Data access and storage
CreateAJxl	Security check department information added to the blockchain	
CreateFJxl	Sorting department information added to the blockchain	
CreateZhuangJxl	Truck loading department information added to the blockchain	
CreateZCxl	Air loading department information added to the blockchain	
CreateDDxl	Arrival department information added to the blockchain	
QueryID	Query the whole process information and the current search times through RFID ID	Data sharing and retrieval
GetbID	Query the whole process information and the current search times through flight number	

Our system developed a consortium blockchain network on the Hyperledger Fabric with nodes which are check-in department node, security check department node, sorting department node, trucking loading department node, air loading department node, and arrival department node. At the same time, we deployed IPFS cluster on the edge nodes. Furthermore, the multichannel architecture was designed to achieve physical isolation for different businesses and coarse-grained access control. The multichannel structure refers to a channel corresponding to a business of airline company or different businesses corresponding to different channels. A channel is parallel to a consortium blockchain. Channels are physically isolated from each other so that ledger information is only visible to the members of the channel thus providing coarse-grained privacy protection. The multichannel design also enables the airport to process multiple types of data and business requirements in parallel and rapidly. Each channel was equipped with smart contract to realize data management and sharing control. Moreover, RFID readers were also deployed on the nodes. Baggage processing information of each department is automatically recorded into the blockchain in real-time through RFID. Conversely, basic baggage information was encrypted by CP-ABE and then written to IPFS deployed at edge nodes. Only the returned address hash was written to the blockchain to enhance the scalability of the blockchain and protect the privacy of passengers. Additionally, smart contract provides interfaces for passengers and airlines, respectively. Passengers and airlines can query baggage tracking information according to RFID ID. In addition, airlines can conduct batch queries through flight number, but only those whose attributes conform to the access control policy can request basic baggage information. At the same time, smart contract calculates baggage requested data for each piece of baggage.

4. CSPBFT Consensus Mechanism

PBFT consensus algorithm is designed to solve the consistency problem of distributed systems with Byzantine nodes [17]. It mainly consists of consistency protocol, view change protocol, and checkpoint protocol, among which consistency protocol is the core. The consistency protocol of PBFT requires to complete two times node communication with

complexity $O(N^2)$, which ensures that the algorithm can achieve consensus even if Byzantine nodes exist in the network. The process of PBFT is mainly shown in Figure 3. Client c is the sender of the request. The primary node receives the requests, sorts them, assigns numbers, and broadcasts them to replicas in the network. The replica is mainly responsible for receiving the messages sent by the primary node and other replicas, carrying out corresponding verification and operations and finally sending the consensus results back to the client [18]. However, there will be a lot of communication between nodes, which will affect the consensus efficiency. In this paper, based on the application scenario of consortium blockchain, we have modified the PBFT algorithm in the following aspects:

4.1. Consistent Protocol Simplification. In the absence of Byzantine nodes, a simplified consistency protocol is adopted to reduce communication traffic between nodes, as shown in Figure 4. The implementation process of the simplified consistency protocol is as follows:

(1) CS-request period

Similar to the request phase of the PBFT algorithm, the client sends a request message to the primary node. The message format is $\langle \text{CS-request}, x, t, c \rangle$, where x is the main content requested by the client, t is the timestamp, and c is the identity information of client C .

(2) CS-preprepare period

After receiving the request message x , the primary node assigns a sequence number n to the received x and then generates a prepreparation message. The message format is $\langle \langle \text{CS-preprepare}, v, n, d, e \rangle, c, x \rangle$. v is the view number, d is the hash calculation result of x , and n is the message number. c is the node integral data, which is used to change the class of the node, and e is the hash calculation result of c . Then, the primary node sends the preprepared messages to all nodes. The consensus node needs to verify the message content. If the verification passes, it will enter the next stage. Otherwise, it will change the view and replace the primary node.

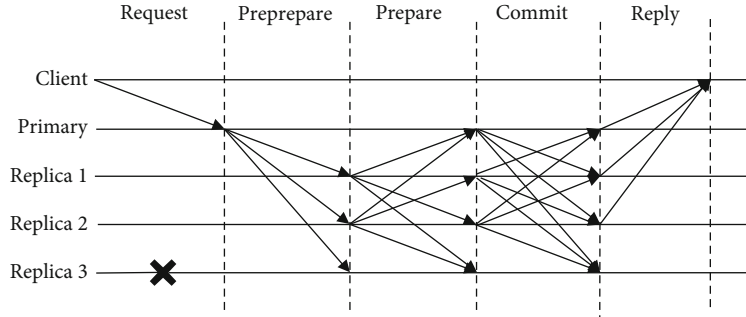


FIGURE 3: PBFT consistency protocol.

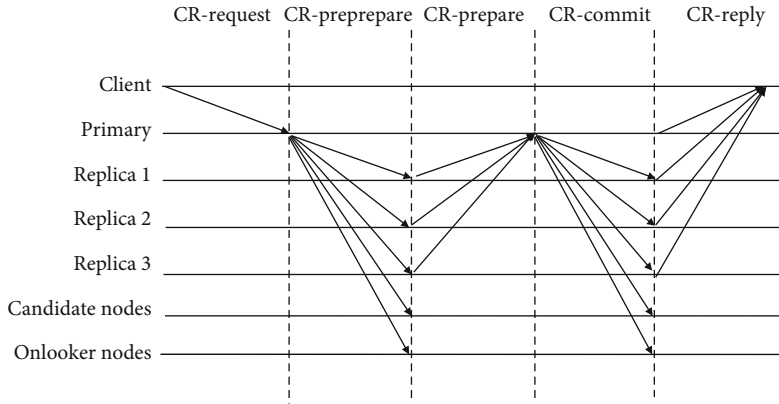


FIGURE 4: Simplified consistency protocol.

(3) CS-prepare period

After verifying the prepared message, the consensus node will judge the information in the verification certificate, including the correctness of blockchain transactions, block header information, and block height. Then, judging whether the integral data in the prepared message is the same as the local integral data, and if not, update the local integral information C . Then, a feedback message is generated and sent to the primary node, which format is $\langle \text{CS-back}, v, n, d, i \rangle$, where i is the number of the node that sending the message.

(4) CS-commit period

If the primary node receives the feedback information sent by $2f + 1$ consensus nodes, and all the feedback information is the same, the primary node will package the feedback information and broadcast it to all nodes in the network. The message format is $\langle \text{CS-commit}, v, n, d, a \rangle$. a indicates that the primary node has confirmed. If the primary node does not receive all the approval information, it will enter the complete consistency protocol process.

(5) CS-response period

Replicas verify whether the approval information of other nodes is correct. If all nodes have received block information, the transaction information will be added to the

local memory. Class C nodes and class D nodes only receive consistent results, but do not give feedback.

4.2. Node Credit Score Mechanism. Selecting the primary nodes in the PBFT algorithm is random. It selects the primary nodes according to number sequence, which is more likely to appear malicious nodes because they are checked. If the malicious nodes are found, the view switching protocol will replace the primary nodes, resulting in a large amount of network communication overhead. In this paper, the PBFT algorithm is improved by node credit score mechanism.

Consortium blockchain requires participants to be authenticated before joining so its credibility and stability are more guaranteed than public blockchain. We apply node credit score mechanism in consortium blockchain. Nodes are selected according to credit scores to ensure higher computing power, wider bandwidth, and stability, which also improved the long-term operating efficiency of the system. With the running of the system, the credit scores and the proportion of malicious nodes in CSPBFT decrease continuously, thus effectively identifying and eliminating malicious nodes as well as ensuring that the algorithm can execute simplified consistency protocol most of the time. The algorithm takes the comprehensive strength of nodes as the initial score basis.

According to the credit scores, nodes are classified into class A, class B, and class C nodes. Class A node has the highest credit rating and takes priority as the primary node.

TABLE 2: Comparison of node permissions.

Credit rating	Taking priority as the primary node	Can act as the primary node	Taking priority as the replicas	Can act as the replicas	Getting consensus results
A	√	√	√	√	√
B	×	√	√	√	√
C	×	×	×	√	√
D	×	×	×	×	√

Secondly, class B nodes participate in consensus as replicas. It also can participate in the election of primary nodes when there is insufficient class A nodes. The node that just joined the system is class C node. The credit rating of class D nodes is too low, so they are not allowed to participate in consensus, but need to accept consensus results. Class C nodes are candidate nodes. When malicious nodes appear in the consensus nodes, the scores of malicious nodes decrease until they are excluded. According to the credit score situation, one node is selected from the class C nodes to join the class B node. If a node successfully participates in block generation, its credit score will increase by 1 point. On the contrary, if a node fails to generate blocks, 5 points of its score will be deducted. When the credit score is lower than 60, a node becomes class D node, which is not allowed to participate in consensus but can accept consensus results. The level of nodes will change in class A, class B, class C, and class D due to their behaviors. Class A and class B nodes are consensus nodes. According to the consistency protocol of the PBFT algorithm, when the client receives more than $f + 1$ consistent messages, it can be considered that the request is successfully executed, so the number of class A nodes is set to $f + 1$ and class B nodes is set to f . Class C nodes, as candidate nodes, do not participate in the system consensus process, whose number is uncertain. Class D nodes, as onlooker nodes, are not allowed to participate in consensus because of their low integral. They only accept consensus results so their number is uncertain. In the experiment, both the number of class C and class D nodes is set to $f/2$. Comparison of node permissions is shown in Table 2.

5. Design and Implementation

As shown in Figure 1, the hybrid architecture has six specific phrases with 14 steps illustrated below. The notations are given in Table 3. It mainly includes the following steps:

5.1. Initialization. At this stage, these entities are initialized: RFID readers, RFID tags, blockchain nodes for different departments, IPFS in edge nodes, and CP-ABE encryption module. Blockchain nodes' access control levels are detailed in Table 4.

The data structure S is defined within a single block. It consists of H_i , public data, and encrypted private data. Public data includes baggage tracking information and baggage request data. Encrypted private data refers to basic baggage information encrypted by CP-ABE. Only the user whose private key completely matches the access control policy can decrypt and obtain passenger information.

TABLE 3: Notations.

Symbol	Description
S	Detail data of a block
P_i	Encryption policy
CT_i	Encrypted basic baggage information
H_i	IPFS address hash
SK_i	Private key of departments in different airlines

TABLE 4: Access control levels of blockchain nodes.

Blockchain nodes	Access	Consensus
Check-in department	Write and read	Yes
Security check department	Write and read	Yes
Sorting department	Write and read	Yes
Trucking loading department	Write and read	Yes
Air loading department	Write and read	Yes
Arrive department	Write and read	Yes
Passenger	Read	No
Administrator	Write and read	Yes

5.2. Encrypt the Basic Baggage Information

Step 1. The CP-ABE encryption module is initialized, and the corresponding private key is assigned according to the attributes of each part of the airport. For example, we set the property of check-in department in airport B to

```
propertyProperty 1 {
    {visitor Sorting 'airline =861202' 'department =05' 'identity =2001192'}
}
```

We set the property of sorting department in airport A to

```
propertyProperty 2 {
    {adminCheckin 'airline =861107' 'department =01' 'identity =1154442'}
}
```

Step 2. Before sharing data, airport A needs to build an encryption policy to achieve access control. We specify the encryption policy p_i which allows airport A to decrypt while airport B cannot as Rule1. After encrypted basic baggage information according to p_i , we get CT_i .

```

rule Rule1 {
    {(admin and (airline<861200 or Checkin))}
    {or (admin and 2 of (department >=01,
identity >=1154442, Sorting))}
}

```

5.3. CT_i Is Uploaded to IPFS in Edge Nodes

Step 3. CT_i is uploaded to the IPFS cluster to ensure not only secure data storage but also colossal storage capacity.

Step 4. IPFS returns the address hash H_i to airport A. CT_i can be queried in IPFS by H_i .

5.4. H_i , RFID ID, and Data Keywords Are Recorded in Blockchain

Step 5. After airport A got H_i from the IPFS cluster, the RFID ID, data keywords, and H_i are uploaded to the blockchain together. Airport A can query the records in the blockchain according to the RFID ID.

Step 6. The consortium blockchain network returns the inquiry result according to the requirements of airport A.

5.5. Baggage Tracking and Complete Baggage Information Are Automatically Formed and Stored in Blockchain

Step 7. Airport A encodes the RFID ID, data keywords, and address hash and stores them in the user area of the RFID tag.

Step 8. Taking six departments as nodes, we build multi-channel consortium chain to realize physical isolation for different companies' businesses as well as coarse-grained privacy protection. The program that automatically stores data in blockchain based on node-serialport, and RFID is deployed on each node. RFID readers were also deployed on the nodes. Node-serialport is a package of Node.js, which is used to read and write serial port data. It is the way to communicate with the RFID reader.

Step 9. When the RFID tags pass through six nodes, there are mainly three steps:

RFID readers receive the data from the user area of the RFID tag, which decodes and intercepts basic baggage information.

The program automatically obtains baggage tracking information, blends it with basic baggage information, and requests to invoke the smart contract.

Smart contract compares data summaries of requests and records in the blockchain. If they are consistent, it allows data to be stored in blockchain.

5.6. Data Sharing and Access Control

Step 10. Airport B can query baggage handling status according to RFID ID or flight number, respectively. Meanwhile, the smart contract will count the baggage tracking request

TABLE 5: System and module versions.

System and module	Versions
Ubuntu	18.04.3 LTS
Hyperledger Fabric	1.3
Docker	19.03.4
Docker-compose	19.03.4
Go	1.12.10
IPFS	0.4.13

data, and the query times of each piece of baggage will be permanently recorded in the blockchain.

Step 11. The consortium blockchain network returns the result to airport B, and airport B can obtain the baggage handling status.

Step 12. In case of lost or damaged baggage, if B demands basic baggage information, it should query its H_i of IPFS in the blockchain according to the RFID ID and then inquires CT_i in IPFS cluster through H_i .

Step 13. The IPFS cluster finds the CT_i and returns query results, and airport B decrypts CT_i according to its private key SK_i . Only when SK_i accord with access policy p can airport B obtain the basic baggage information.

Step 14. Passengers can query real-time baggage tracking information according to the RFID ID. Meanwhile, the query times of each piece of baggage will also be recorded in the blockchain. Administrators can conduct not only batch or accurate retrieval, but also data interaction.

6. Experiments and Evaluation

6.1. Experimental Setup. In this section, we implement experiments to evaluate the performance of the proposed hybrid baggage tracking system that was prototyped on the Hyperledger Fabric. The specific configuration of the experimental platform and the experimental environment is as follows: the system is deployed on 2 hosts with intel corei7-9700@3.00 GHz processor and corei7-5500@2.40 GHz processor, 4 GB RAM, and we have two RFID readers. The system and module versions are shown in Table 5.

6.2. Experimental Results. (1) *Query Real-Time Baggage Tracking Information According to the RFID ID.* In the first experiment, we log into the blockchain network as administrators. Not only can we conduct bulk queries according to flight number but also accurate queries by RFID ID. We can get the time of baggage arrival in each department, flight number, and IPFS address hash corresponding to basic baggage information, as shown in Figure 5. At the same time, this retrieval behavior will be permanently recorded by blockchain network, and the number of queries will plus one.

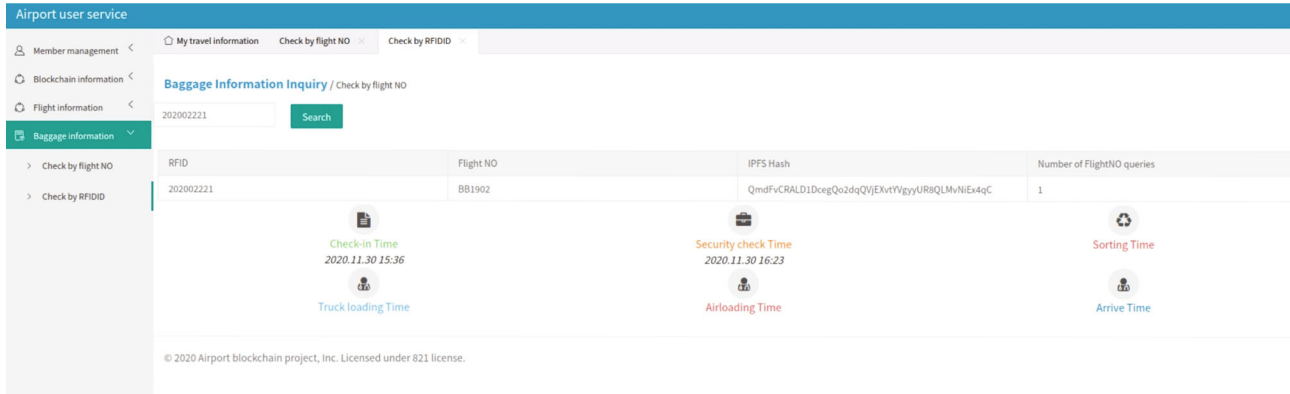


FIGURE 5: Query real-time baggage tracking information by RFID ID.

```
wangfei@wangfei:~/wish$ cpabe-dec pub_key AirportBSorting_priv_key PassengerInformation.js.cpabe
cannot decrypt, attributes in key do not satisfy policy
```

FIGURE 6: The unauthorized airport department could not decrypt.

(2) *Test of Access Control.* In the second experiment, only when the department's property conforms to the access control policy can it decrypt basic baggage information downloaded from IPFS. For instance, the access policy we developed in attribute-based access control allows all departments in airport A to decrypt basic baggage information while airport B cannot, as shown in Figure 6.

(3) *System Performance.* We conducted four rounds of tests to measure the average transaction processing time, the average time for RFID data to store in blockchain, and the average time to reject retrieval that does not comply with the access control policy. To evaluate the scalability of our system, we also carry out extensive system performance evaluations by increasing the number of baggage and departments. The average transaction processing time from 4 rounds of tests remains at around 0.40s when the number of departments increased from 4 to 20 as shown in Figure 7. The average time to reject retrieval that does not comply with the access control policy from 4 rounds of tests remains at around 50 ms as shown in Figure 8. The average time for RFID data to store in the blockchain from 4 rounds of tests remains at around 0.58 s when the number of baggage increased from 6 to 1000 as shown in Figure 9. The results of the experiments show that when the number of baggage and departments increased, transaction time and response time did not change significantly, which demonstrates that the performance of the proposed system is scalable. The prototype system can realize baggage processing information stored in the blockchain in real time and automatically as well as produce response in a few hundred milliseconds, which makes it suitable in practical baggage tracking systems.

6.3. *CSPBFT Performance.* In this section, we compare PBFT and CSPBFT consensus algorithms in terms of communication latency, communication overhead, and operational efficiency through experiments. This experiment simulates a multinode consortium blockchain system by JAVA.

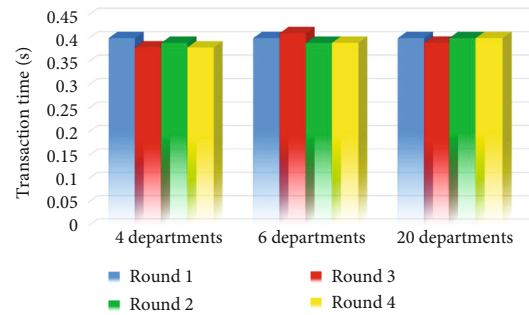


FIGURE 7: The average transaction processing time.

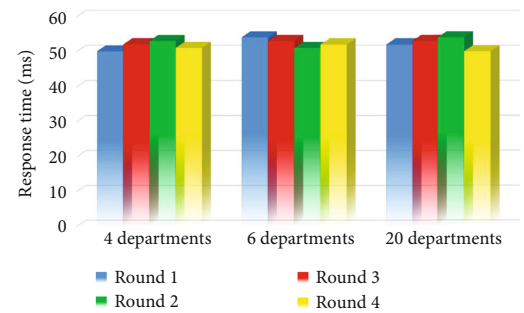


FIGURE 8: The average time to reject retrieval that does not comply with the access control policy.

- (1) *Communication Latency.* Communication latency refers to the time interval between the client sending a transaction request to the primary node and the client confirming the completion of consensus, which is an essential parameter for evaluating the performance of consensus algorithm. Reducing communication latency can improve the efficiency and practicability of the system. In this experiment, the total number of nodes in the system is taken as an

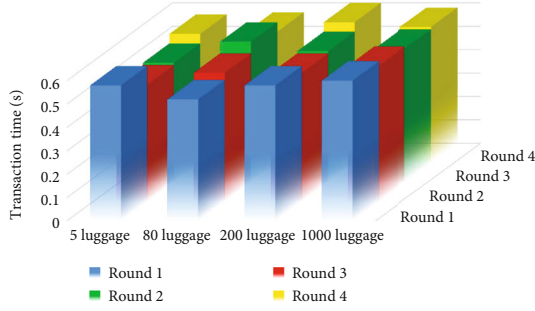


FIGURE 9: The average time for RFID data to store in blockchain.

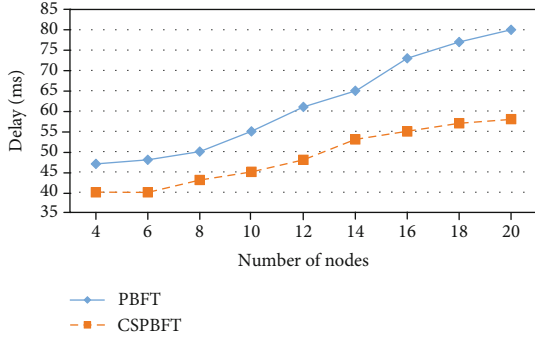


FIGURE 10: Communication latency contrast without Byzantine nodes.

experimental variable. The number of nodes increases from 4 to 20. The step size is 2. Transactions are carried out under different node numbers, and the average value in different states is taken as the final value of the communication latency. As shown in Figure 10, in the absence of Byzantine nodes, the CSPBFT algorithm implements a simplified consistency protocol, which is superior to the PBFT algorithm in communication latency. With the increase of the number of nodes, the CSPBFT algorithm has lower communication latency growth rate and better stability. However, in the presence of Byzantine nodes, the communication latency of the CSPBFT algorithm increases obviously due to the switching of consensus protocols, as shown in Figure 11. However, through the node credit score mechanism, the fault nodes in the system can be effectively identified and eliminated, so that the algorithm can execute the simplified consistency protocol most time, thus improving the long-term operational efficiency of the system.

- (2) *Communication Overhead.* The PBFT algorithm has three core stages, which are preprepare stage, prepare stage, and commit stage. The maximum tolerance number of fault nodes in the blockchain is f , and the number of nodes i in the system should not be less than $3f + 1$. Considering the consensus

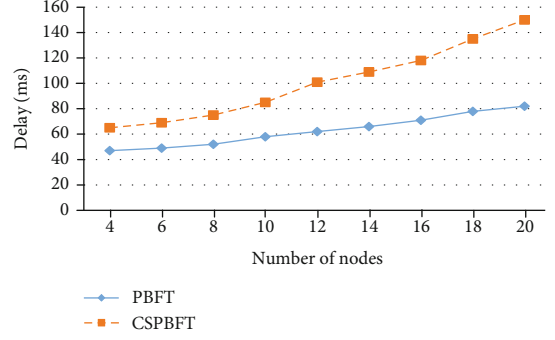


FIGURE 11: Communication latency contrast when Byzantine nodes exist.

efficiency of the system, i is usually $3f + 1$. Assuming that all nodes in the network communicate normally, the calculation of total message volume in the three-stage consensus process of PBFT is shown in Formula (5).

$$M_{\text{pbft}} = 6f(3f + 1). \quad (5)$$

As for CSPBFT, in the main four-stage consensus process of preprepare stage, prepare stage, commit stage, and response stage, when all nodes in the network communicate normally, the calculation of message volume to generate a new block is as follows.

$$M_{\text{cspbft}} = 9f. \quad (6)$$

When the primary node fails and needs view conversion, the replicas need to communicate view conversion information through pairwise interaction. For PBFT, the communication amount is $9f^2$ at this time. After completing the view change, the primary node needs to send a view confirmation message to the replicas, and the communication amount is $3f$. Combined with the view change probability p , the average total communication amount of the PBFT algorithm is as follows:

$$C_{\text{pbft}} = 18f^2 + 6f + p(9f^2 + 3f). \quad (7)$$

For CSPBFT, the average total communication amount is

$$C_{\text{cspbft}} = 9f + p(4f^2 + 2f). \quad (8)$$

Therefore, the ratio Q of communication amount between CSPBFT and PBFT is as follows:

$$Q = 18f^2 + 6f + p(9f^2 + 3f) \{9f + p(4f^2 + 2f)\}. \quad (9)$$

The visual graph of Q is obtained by MATLAB. The value of p ranges from 0 to 1, the step size is 0.1, the value of f ranges from 3 to 33, and the step size is 3, as shown in Figure 12.

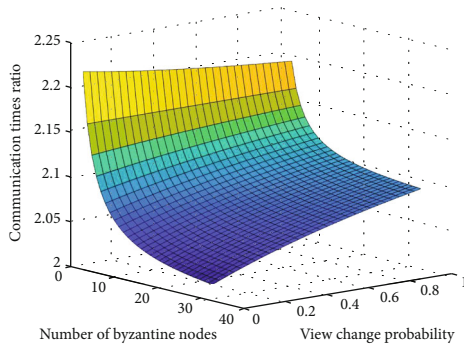


FIGURE 12: Communication overhead contrast.

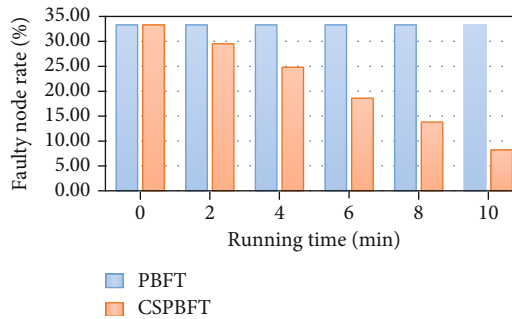


FIGURE 13: Faulty node rate.

It can be seen from the figure that no matter how the values of p and f change, the value of Q is always less than 1. This suggests that the communication volume of CSPBFT is always less than PBFT. With the increase of nodes, the Q decreases gradually, which indicates that the communication overhead performance of the CSPBFT algorithm is still better than PBFT in multinode environment. In addition, the CSPBFT algorithm introduces node credit score mechanism to evaluate node behavior, which reduces the probability of Byzantine node. Therefore, the CSPBFT algorithm has better performance in communication overhead in the practical process.

- (3) *Operating Efficiency.* One of the purposes of CSPBFT design is to improve the long-term operational efficiency of the system. With the running of the system, the credit scores and the proportion of malicious nodes in CSPBFT decrease continuously, thus effectively identifying and eliminating malicious nodes in the system. However, the faulty node rate in PBFT does not change. Figure 13 shows the change of the faulty node rate between PBFT and CSPBFT for a long time. Therefore, through simplified consistency protocol, CSPBFT can generate blocks more efficiently than PBFT.

7. Conclusion

In this paper, we have proposed a system named BEI-TAB that utilized RFID combined with consortium blockchain to realize the real-time tracking information automatically stored in the blockchain, which not only avoids data leakage

but also improves the industrialization level of the airport. In addition, we took advantage of multichannel architecture realized physical isolation of different businesses and coarse-grained privacy protection. At the same time, we utilized CP-ABE and IPFS to store basic baggage information in edge nodes so as to improve the scalability of blockchain and provide fine-grained privacy protection. To this end, we have deployed a testbed with both software and hardware to evaluate the performance of transaction processing time and speed. The experiments showed that our system is scalable, which makes it suitable to be incorporated in secured and real-time baggage tracking. Besides, we improved the PBFT algorithm to CSPBFT which adopted a simplified consistency protocol to reduce the communication traffic between nodes in the absence of Byzantine nodes. The Byzantine nodes in the system are effectively identified and excluded by the node credit score mechanism, so that the algorithm can execute the simplified consistency protocol most time, thus improving the long-term operational efficiency of the system.

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] IATA, "Resolution: RFID baggage tracking set for global deployment," <https://www.iata.org/en/pressroom/pr/2019-06-02-05/>.
- [2] H. Ding, X. Li, Y. Cai, B. Lorenzo, and Y. Fang, "Intelligent data transportation in smart cities: a spectrum-aware approach," *IEEE/ACM Transactions on Networking*, vol. 26, no. 6, pp. 2598–2611, 2018.
- [3] H. Ding, C. Zhang, Y. Cai, and Y. Fang, "Smart cities on wheels: a newly emerging vehicular cognitive capability harvesting network for data transportation," *IEEE Wireless Communications*, vol. 25, no. 2, pp. 160–169, 2018.
- [4] T. Ahmed, T. Calders, and T. B. Pedersen, "Mining risk factors in RFID baggage tracking data," in *2015 16th IEEE International Conference on Mobile Data Management*, pp. 235–242, Pittsburgh, PA, USA, June 2015.
- [5] R. Yang, F. R. Yu, P. Si, Z. Yang, and Y. Zhang, "Integrated blockchain and edge computing systems: a survey, some research issues and challenges," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1508–1532, 2019.
- [6] H. Huang, J. Lin, B. Zheng, Z. Zheng, and J. Bian, "When blockchain meets distributed file systems: an overview, challenges and open issues," *IEEE Access*, vol. 8, pp. 50574–50586, 2020.
- [7] A. Singh, S. Meshram, T. Gujar, and P. R. Wankhede, "Baggage tracing and handling system using RFID and IoT for airports," in *2016 International Conference on Computing, Analytics and Security Trends (CAST)*, pp. 466–470, Pune, India, December 2016.

- [8] M. S. Jerry, M. M. Vijay, and T. K. Tulshiram, "Carousel security management and cargo deck tracking of passenger baggage using wireless technology," in *2016 IEEE Bombay Section Symposium (IBSS)*, pp. 1–6, Baramati, India, December 2016.
- [9] M. Johnson and A. Gilman, "Real-time baggage tracking using a modified background subtraction algorithm," in *2012 19th International Conference on Mechatronics and Machine Vision in Practice (M2VIP)*, pp. 200–204, Auckland, New Zealand, 2012.
- [10] Q. Gao and P. Liang, "Airline baggage appearance transportability detection based on a novel dataset and sequential hierarchical sampling CNN model," *IEEE Access*, vol. 9, pp. 41833–41843, 2021.
- [11] S. Wang, Y. Zhang, and Y. Zhang, "A blockchain-based framework for data sharing with fine-grained access control in decentralized storage systems," *IEEE Access*, vol. 6, pp. 38437–38450, 2018.
- [12] K. Christidis and M. Devetsikiotis, "Blockchains and smart contracts for the Internet of Things," *IEEE Access*, vol. 4, pp. 2292–2303, 2016.
- [13] A. Juels, "RFID security and privacy: a research survey," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 2, pp. 381–394, 2006.
- [14] R. Norvill, B. B. FizPontiveros, R. State, and A. Cullen, "IPFS for reduction of chain size in Ethereum," in *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, pp. 1121–1128, Halifax, NS, Canada, 2018.
- [15] H. Ding, Y. Guo, X. Li, and Y. Fang, "Beef up the edge: spectrum-aware placement of edge computing services for the Internet of Things," *IEEE Transactions on Mobile Computing*, vol. 18, no. 12, pp. 2783–2795, 2019.
- [16] J. Bethencourt, A. Sahai, and B. Waters, "Ciphertext-policy attribute-based encryption," in *2007 IEEE Symposium on Security and Privacy (SP '07)*, pp. 321–334, Berkeley, CA, USA, May 2007.
- [17] K. Lei, Q. Zhang, L. Xu, and Z. Qi, "Reputation-based Byzantine fault-tolerance for consortium blockchain," in *2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS)*, pp. 604–611, Singapore, December 2018.
- [18] G. S. Veronese, M. Correia, A. N. Bessani, L. C. Lung, and P. Verissimo, "Efficient Byzantine fault-tolerance," *IEEE Transactions on Computers*, vol. 62, no. 1, pp. 16–30, 2013.

Research Article

A Weakly Supervised Academic Search Model Based on Knowledge-Enhanced Feature Representation

Mingying Xu , Junping Du , Feifei Kou , Meiyu Liang , Xin Xu , and Jiaxin Yang 

Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia, School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China

Correspondence should be addressed to Junping Du; junpingdu@126.com

Received 11 May 2021; Revised 30 July 2021; Accepted 11 August 2021; Published 22 September 2021

Academic Editor: Bo Rong

Copyright © 2021 Mingying 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.

Internet of Things search has great potential applications with the rapid development of Internet of Things technology. Combining Internet of Things technology and academic search to build academic search framework based on Internet of Things is an effective solution to realize massive academic resource search. Recently, the academic big data has been characterized by a large number of types and spanning many fields. The traditional web search technology is no longer suitable for the search environment of academic big data. Thus, this paper designs academic search framework based on Internet of Things Technology. In order to alleviate the pressure of the cloud server processing massive academic big data, the edge server is introduced to clean and remove the redundancy of the data to form a clean data for further analysis and processing by the cloud server. Edge computing network effectively makes up for the deficiency of cloud computing in the conditions of distributed and high concurrent access, reduces long-distance data transmission, and improves the quality of network user experience. For Academic Search, this paper proposes a novel weakly supervised academic search model based on knowledge-enhanced feature representation. The proposed model can relieve high cost of acquisition of manually labeled data by obtaining a lot of pseudolabeled data and consider word-level interactive matching and sentence-level semantic matching for more accurate matching in the process of academic search. The experimental result on academic datasets demonstrate that the performance of the proposed model is much better than that of the existing methods.

1. Introduction

Internet of Things technology has contributed to the realization of many Internet of Things applications that benefit the whole world and constantly changes people's way of life. Internet of Things search service is one of the most important services provided by the Internet of Things. It can efficiently and accurately obtain information to meet the needs of users from massive, heterogeneous, and dynamic Internet of Things data. Currently, academic big data presents the characteristics of huge quantity, various types, and spanning multiple fields. Different from network data in the general sense, academic big data includes paper patents, scientific research equipment, experts and scholars, scientific research teams, national key laboratories, major scientific research infrastructure and large scientific research

instruments, academic video scientific research reports, and other scientific and technological entities. Academic big data presents the characteristics of massive, multisource, heterogeneous, and related. Traditional web search technology is no longer suitable for academic big data search environment. Internet of Things search can solve the search problem of academic big data. Internet of Things search is that users send search query requests to the network system, the network system exchanges information with the physical world, and then returns the search object and its location, status, and other information to users. Academic big data search based on Internet of Things technology transfers the massive academic big data to remote cloud server for analysis and calculation, which promotes the interaction between users and intelligent system, improves the search efficiency and realizes automation of academic search. Academic

search should not only ensure the search efficiency but also ensure the search accuracy. Therefore, efficient and accurate search is urgently needed.

However, on the one hand, there are several key issues to be considered in the way that massive academic big data is analyzed and calculated by remote cloud server. First of all, multisource, heterogeneous massive big data has data redundancy, noise, data missing, and other problems. If it is transmitted to the remote cloud processing server, it will increase the burden of the cloud server. Secondly, a large number of data directly communicate with the cloud server, which occupies the bandwidth resources and seriously affects the transmission rate. At the same time, accurate academic resource search is another problem to be solved. Many current academic search studies pay little attention on specific search algorithms [1, 2]. There are only few studies on specific academic search task. Explicit Semantic Ranking [3] (ESR) defines academic retrieval as entity set retrieval, which represents the query and each document using knowledge graph embedding, and uses manually labeled training data to train a supervised academic search model. SetRank [4] models the relationships between entities through the type of entities while representing entities and proposes an unsupervised academic search framework. The retrieval performance of ESR and SetRank based on entity set retrieval has been improved to some extent, but there is still much room for improvement. On the one hand, ESR does not consider the relationship between entities. On the other hand, ESR only considers entity level matching, but does not consider deep semantic matching between query and text. What is more, ESR is a supervised academic retrieval model. It can show strong effectiveness [5] when large-scale manual-labeled data of document relevance are available. But manually labeled academic data usually requires domain expert knowledge, which is time-consuming, labor-intensive, and difficult to obtain. It is a serious bottleneck [6] for supervised academic retrieval. Moreover, although SetRank models interentity relationships using entity types, it still cannot model complex relationships between entities. In many cases, a user who submits such a query as “a new ranking method that leverages knowledge graph embedding” will expect to know how “ranking method” associated with “knowledge graph embedding.” The distinguishing characteristic of such queries is that they reflect user’s needs to find documents containing interentity relationships. As the results of the survey in [4], such queries are common in academic search scenarios.

In response to the above-mentioned problems, this paper designs academic search framework in technology Internet of Things. In order to alleviate the pressure of the cloud server processing massive academic big data, the edge server is introduced to clean and remove the redundancy of the data to form a clean data for further analysis and processing by the cloud server. The edge computing network effectively makes up for the deficiency of cloud computing in the conditions of distributed and high concurrent access, reduces the long-distance data transmission process, has faster processing and response speed and lower computing and storage costs, and greatly improves the quality of network

user experience. For academic search, this paper proposes a weakly supervised academic search model based on knowledge-enhanced feature representation. In this paper, we refer to both entities and words as features. Specifically, we first employ a scientific information organization tool SCIE [7] to extract entities and relationships from scientific literatures. In order to obtain the feature representation that can express the entity and the relationship between entities, language model is trained with structured knowledge added as supervision signals. Based on the learned feature vector, we propose a weakly supervised academic retrieval model. In academic search, we should not only pay attention to the semantic matching of words in the query and document but also consider interactive matching between the features in the query and the document. On the other hand, a new weakly supervised method suitable for academic search is employed to obtain a large amount of pseudolabeled training data. Then, the academic search model is trained on labeled training data and pseudolabeled training data based on the knowledge-enhanced feature representation. We conduct extensive experiments on academic data sets. The experimental results demonstrate the effectiveness of our model in improving retrieval performance. We summarize the main novelties and contributions as follows:

- (1) An academic search framework based on Internet of Things technology is designed
- (2) A novel weakly supervised academic search model based on the knowledge-enhanced feature representation is proposed to improve academic search performance
- (3) In the process of academic search, not only the word-level interactive matching but also the sentence-level semantic matching between query and document are considered to realize the accurate matching between query and document
- (4) Extensive experiments on the academic datasets prove that our proposed model is significantly better than the state-of-the-art search methods

The rest of this paper is organized as follows: Section 2 discusses related work, and Section 3 describes knowledge-enhanced feature representation for weakly supervised academic search. Section 4 reports experimental results and analysis, and we summarize this article in Section 5.

2. Related Work

2.1. IOT Search. Internet of Things search service is that users send search requests to the network system, the network system exchanges information with the physical world, and then returns information such as the location and status of the search object to users, which is directly driven by users. The existing Internet of Things search technology mainly includes location-based search, content-based search, and heterogeneous search. Location-based search mainly searches the content associated with location. The location

information may be expressed as geographic coordinates or may be a logical location, such as a distance from another device [8]. Content-based search is based on the data content collected by a specific target sensor. First, the Internet of Things search engine analyzes the content and maps the corresponding index. Then, when querying, the search engine uses the corresponding index to pair the query with the content and returns the sensor information. Heterogeneous search technology mainly includes semantic or ontology-based search and resource retrieval. In particular, ontology represents concepts, types, and relationships in different fields [9]. The integration of semantic and ontology mechanism can help the system to build domain, task, and method combination search system. On the other hand, all data and IOT devices can abstract resources and provide different services.

2.2. Knowledge-Aware Representation. Representation learning [10] is a core issue in the information retrieval field. Distributed representation methods such as word2vec [11] and glove [12] have been successfully applied in the field of information retrieval [13]. In recent years, the emergence of large-scale knowledge graphs has promoted the development of knowledge-aware representation [14]. The knowledge graph contains rich knowledge [15]. Knowledge-aware representation brings in rich semantic information from the knowledge graph, which significantly improves the effectiveness of the search algorithms [16] and provides new opportunities for better understanding queries and documents [17, 18]. For instance, Xiong et al. introduce a bag-of-entities model, which successfully improves the retrieval accuracy by representing queries and documents using their entity annotations [19]. Hadas et al. design an entity-based language model which uniformly marks the individual features in the text and the term sequence recognized as entities by the entity linking tool as entities and effectively used for document retrieval [20]. A word-entity framework [21] is proposed that combines the bag-of-words and the entities linked to the knowledge graph to optimize information retrieval. Liu et al. employ Entity-Duet Neural Ranking [22], which introduces the knowledge graph into the neural search system. This model proposes a knowledge representation method combining words and entities to represent queries and documents and significantly improves search performance.

2.3. Weakly Supervised Neural Information Retrieval. With the great success of deep neural networks, several deep neural network-based retrieval models have been proposed, such as CDSSM [23], MatchPramid [24], DRMM [25], K-NRM [26], and CONV-KNRM [27]. These methods focus on either matching the whole document or word level interaction. There is no balance between the two aspects. What is more, when large-scale relevance training signals are available, the neural network retrieval models have shown strong effectiveness. However, collecting manually labeled data is time-consuming and labor-intensive [28]. Therefore, insufficient labeled training data has become the main obstacle that affects the performance of neural network-based retrieval models [29]. Weak supervision [30, 31] is an effective way

to overcome this limitation. Levy et al. use the generated pseudo-labeled data to train the neural network retrieval model, thereby solving the problem of difficulty in obtaining labeled data [32]. One can obtain clicked data of users as weakly supervised training data [33–35]. However, the randomness and subjectivity of user's click behavior caused the inconsistency between user's click and the real relevance label, which made the training set mixed with noise. Dehghani et al. propose a ranking-based method [36] to train a neural retrieval model with the help of an existing retrieval model such as BM25, which assumes that documents with higher scores are more relevant to the query than documents with lower scores. Recently, content-based methods [37, 38] are proposed to generate a set of weak pseudoqueries and related documents. MacAvaney et al. employ filters to eliminate training samples that cannot be converted well into relevance scores [37]. A training sample filtering technique based on heuristics and a new type of supervised filtering [38] is developed to remove those samples which are far away from the domains.

3. Weakly Supervised Academic Search Based on Knowledge-Enhanced Feature Representation in Technology Internet of Things

3.1. System Model. In this part, we introduce the architecture of academic search system in academic Internet of Things. As shown in Figure 1, the system is mainly divided into four layers, data acquisition layer, edge server layer, cloud server layer, and academic search layer.

3.1.1. Data Acquisition. The data acquisition layer mainly obtains academic big data resources, including papers and patents, scientific research teams, scientific and technological talents, key laboratories, instruments, and equipment.

3.1.2. Edge Server. Edge server adopts edge computing model, which is a distributed network model. Because of its high autonomy, each edge server is located in a specific area and connected to a specific device or data source in that area. The edge server is used to clean the academic resources, remove the redundancy, and submit them to the cloud server for further analysis and processing.

3.1.3. Cloud Server. The cloud server layer is composed of a large number of dedicated servers with strong computing power, storage capacity, and stable connection, which can process a large number of data and complex computing in a very short time. Therefore, the most important advantage of the cloud server layer is the aggregation, mining, analysis [39, 40], and storage of massive data, which are beyond the processing capacity of the edge server. The cloud server layer is mainly used to deal with more complex and huge computing intensive data and tasks, such as the training of academic search model based on deep learning.

3.1.4. Academic Search Layer. According to the query submitted by users, it aims to search a group of academic Internet of Things resources, including academic resources and

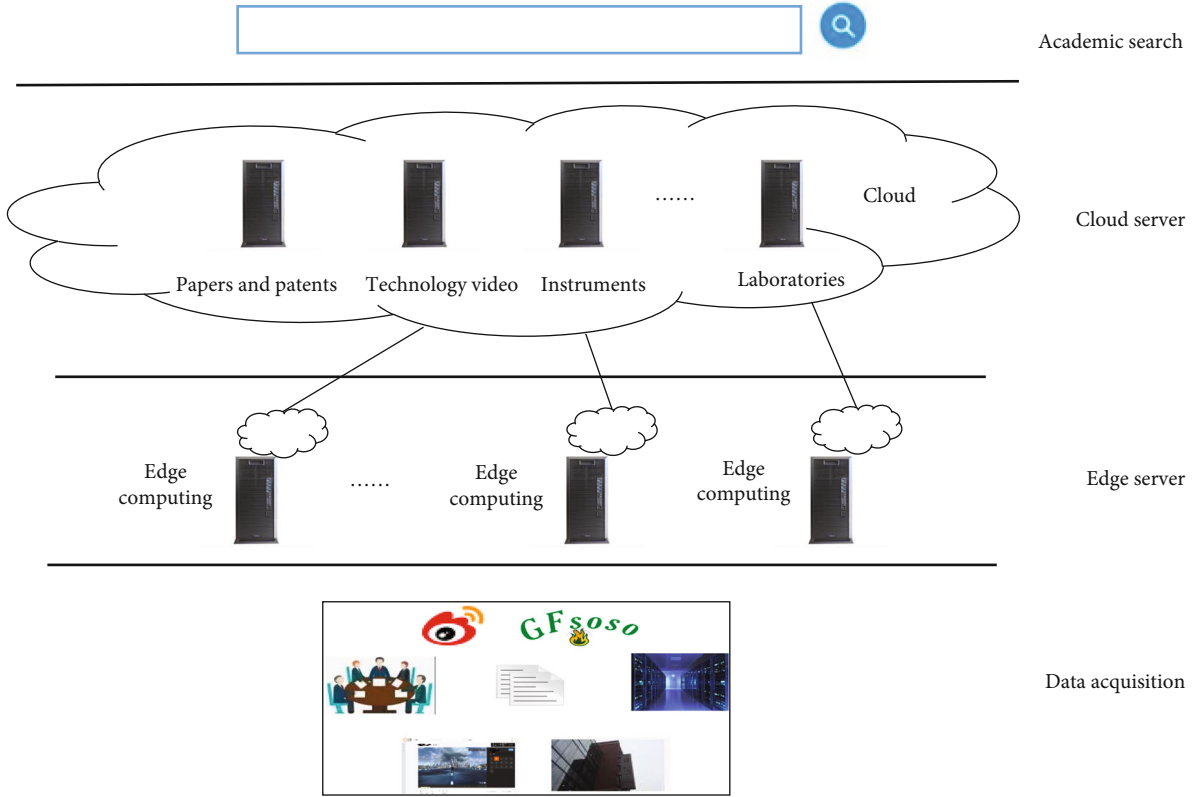


FIGURE 1: Academic Internet of Things framework.

equipment. The Internet of Things search engine responds to the query and returns academic resources (such as key laboratories and equipment), academic data (such as papers and patents).

The academic search process is shown in Figure 2 as follows.

3.2. Weakly Supervised Academic Search Based on Knowledge-Enhanced Representation. Weakly supervised academic search model is shown in Figure 3, including two main processes: knowledge-enhanced feature learning and weakly supervised academic search.

For the knowledge-enhanced feature learning, it models entities and their relationships by adding the knowledge extracted from scientific literatures to train word2vec to express rich semantics. The training process makes the learned features not only depend on the cooccurrence information of features in context but also learn the complex relationship between features.

Weakly supervised academic search model can be trained on large-scale data with the help of weakly supervision strategy to improve the search performance. In the process of academic search, the representation-based matching model and the interaction-based matching model are deployed to the matching module. The model focuses on not only the accurate matching between words but also the semantic matching between query and document to improve the matching accuracy.

3.2.1. Knowledge-Enhanced Feature Representation Learning. In this paper, we propose a knowledge-enhanced feature representation learning method by using word2vec based on skip-gram. The proposed representation learning method is able to express richer semantic information with the knowledge graph embedding model TransE [41]. This paper uses the scientific information extraction tool SCIIE to extract the entities and the relationships in the academic literature. SCIIE defines 6 relationships including “Used-for,” “Feature-of,” “Hyponym-of,” “Part-of,” “Compare,” and “Conjunction.” It can extract entity and interentity relationships in scientific papers and organize them into structured knowledge $T(t_i, r, t_j)$. Here, t_i and t_j are features extracted from the scientific literature, and r is relationship between t_i and t_j . This knowledge is added when training word2vec, so that the learning process of feature vectors is based on not only the cooccurrence information of the context but also the relationship between features, thereby improving the quality of semantic expression. Here, we refer to word and entity extracted from the training corpus as features. The objective function of the knowledge-enhanced feature representation learning model is as follows:

$$L = \sum_i \log p(t_c | t_i) + \sum_r \log p(t_j | t_i + r), \quad (1)$$

where t_i is a feature of the entire corpus and t_c is the

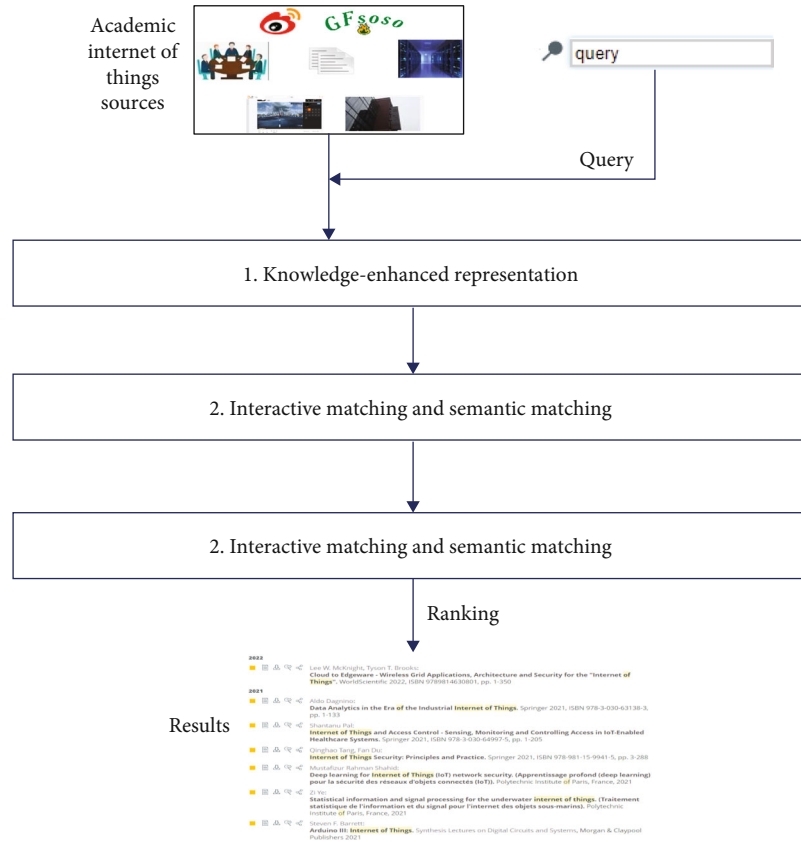


FIGURE 2: Academic search process.

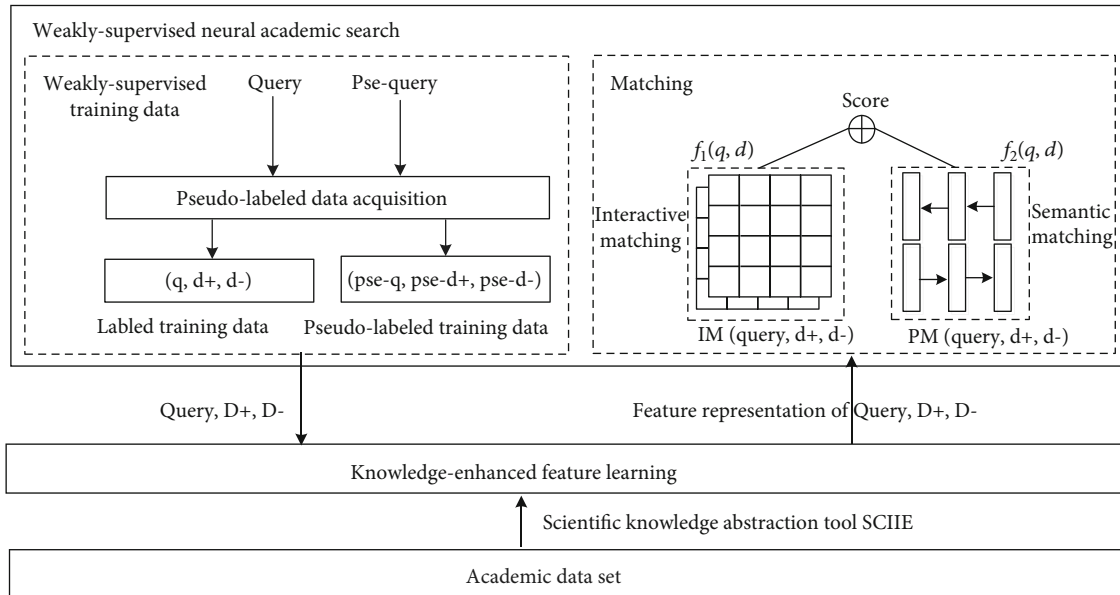


FIGURE 3: Weakly supervised academic search model.

context of t_i , r is the relationship between t_i and t_j . The left side of Equation (1) is the optimization loss of word2vec based on skip-gram, and the right side is the optimization

loss of TransE. $p(t_c | t_i)$ is the probability that t_c be predicted as the context of term t_i . $p(t_j | t_i + r)$ is the probability that $t_j + r$ equals to t_j . They can be approximately calculated by

negative sampling. Given t_i and r , t_c and t_j are specifically positive samples of word2vec model and TransE model. We use negative sampling to extract negative samples of t_i .

The optimization loss of word2vec based on skip-gram can be calculated as follows:

$$\sum_i \log p(t_c | t_i) = \sum_{(t_i, t_c) \in S^+} \log \sigma(e_{t_c} \bullet e_{t_i}) + \sum_{t_{c'} \in N(t_i)} \log \sigma(e_{t_{c'}} \bullet e_{t_i}). \quad (2)$$

Here, $(t_i, t_c) \in S^+$ implies that t_i is the context of t_c , S^+ is positive sample sets of t_i , $t_{c'} \in N(t_i)$ implies that $t_{c'}$ is not the context of t_i , $N(t_i)$ is the negative sample set of t_i , e_{t_i} , e_{t_c} , $e_{t_{c'}}$ is the embedding of feature t_i , t_c , $t_{c'}$.

The optimization loss of TransE can be calculated as follows:

$$\begin{aligned} \sum_r \log p(t_j | t_i + r) &= \sum_{(t_i, r, t_j) \in T^+} \log \sigma(e_{t_i+r} \bullet e_{t_j}) \\ &+ \sum_{(t_i, r, t_j) \in T^-} \log (1 - \sigma(e_{t_i+r} \bullet e_{t_j})). \end{aligned} \quad (3)$$

$(t_i, r, t_j) \in T^+$ implies that $t_i + r$ equals to t_j . T^+ is the positive sample set of (t_i, r) . $(t_i, r, t_j) \in T^-$ implies that $t_i + r$ do not equal to t_j . T^- is the negative sample set of (t_i, r) .

We use stochastic gradient ascent to update parameters. Finally, feature vectors e_{t_i} can be obtained which express semantic relations among entities and entity relationship.

3.2.2. Weakly Supervised Academic Search. The weakly supervised academic search model is implemented as a pairwise ranking model in Figure 1. Given a query-document pair (q, d_1, d_2) , it tries to learn a model that assigns a larger score to document d_1 than document d_2 if d_1 matches to q better. By combining the ranking-based weak supervision method with the content-based weak supervision method, we propose a novel weak supervision method for generating query document pairs.

On the one hand, for a given query and of manually labeled documents, a query-document pair is formed as labeled training data (query, $d+$, $d-$). On the other hand, given a query, top-ranked documents can be recalled from candidate documents using BM25 algorithm. Then, human-labeled documents related to the query are labeled as positive samples, and the documents recalled by BM25 are labeled as negative samples, which forms positive and negative sample pairs. At the same time, considering the title of the paper reflects the most critical content of the paper in the most appropriate and concise terms. And the abstract of the paper contains the most important and necessary information of the paper and is a summary of the main content of the paper. So, we choose the title and corresponding abstract of a paper in the document collection as a positive query-document sample. To sample a negative example of the query, we still use the BM25 algorithm to search the corresponding negative example of this query. Then, the pseudopositive and negative

sample pairs are formed, denoted by Pse-training data (Pse-query, Pse-D+, Pse-D-). Finally, weakly supervised training data (Query, $D+$, $D-$) are formed including labeled training data (query, $d+$, $d-$) and pseudolabeled training data (pse-query, pse-d+, pse-d-). The weakly supervised training data are sent to knowledge-enhanced representation layer to map query and document to feature embedding $\{e_t^q\}_{t=1}^m$ and $\{e_t^{id}\}_{t=1}^n$. Then, $\{e_t^q\}_{t=1}^m$ and $\{e_t^{id}\}_{t=1}^n$ are sent to matching layer to calculate matching score1 and score2. Score1 represents the interaction matching score of the word in query and document, and score2 represents the semantic matching score of query and document.

Specifically, KNRM [26] is employed to compute interactive matching score score1. First, the transformation matrix M_{ij} calculates the similarity between the word of query and document.

$$M_{ij} = \cos(e_i^q, e_j^d). \quad (4)$$

Then, kernels are employed to convert M to query-document matching features $\emptyset(M)$.

$$\begin{aligned} K_K(M_i) &= \sum_j \exp\left(-\frac{(M_{ij} - \mu_k)^2}{2\sigma_k^2}\right), \\ \vec{K}(M_i) &= \{K_1(M_i), \dots, K_K(M_i)\}, \\ \emptyset(M) &= \sum_{i=1}^n \log \vec{K}(M_i). \end{aligned} \quad (5)$$

$K_K(M_i)$ represents the k^{th} RBF kernel. $\vec{K}(M_i)$ employs K kernels to transform translation matrix into a K -dimensional feature vector. The matching features $\emptyset(M)$ are sent to a fully connected layer to produce the final ranking score $f_1(q, d)$.

$$f_1(q, d) = \tanh\left(W_{s_1}^T \emptyset(M) + b_{s_1}\right). \quad (6)$$

At the same time, $\{e_t^q\}_{t=1}^m$ and $\{e_t^{id}\}_{t=1}^n$ are sent to a simple Bi-LSTM layer, and we can get the contextual representation. The specific calculation is as follows:

$$\begin{aligned} c_t^q &= \text{BiLSTM}_q(c_{t-1}^q, e_t^q), \\ c_t^{id} &= \text{BiLSTM}_q(c_{t-1}^{id}, e_t^{id}). \end{aligned} \quad (7)$$

Then, self-attention is employed to distinguish the importance of different words in the query and sentence of document when calculating the representation of query and document.

$$\begin{aligned}
a_t^q &= \frac{\exp\left(W_q^T(\tanh(W_q \bullet c_t^q))\right)}{\sum_{i=1}^m \exp\left(W_q^T(\tanh(W_q \bullet c_i^q))\right)}, \\
\text{pre}^q &= \sum_{j=1}^m a_j^q c_j^q, \\
a_t^{id} &= \frac{\exp\left(W_d^T(\tanh(W_d \bullet c_t^{id}))\right)}{\sum_{j=1}^n \exp\left(W_d^T(\tanh(W_d \bullet c_j^{id}))\right)}, \\
\text{pre}^{id} &= \sum_{t=1}^n a_t^{id} c_t^{id}, \\
\text{pre}^d &= \frac{\sum_{t=1}^l \text{pre}^{td}}{l}.
\end{aligned} \tag{8}$$

Finally, the similarity score $f_2(q, d)$ of query and document is calculated through a full connection layer.

$$f_2(q, d) = \tanh\left(W_{s_2}^T(\text{pre}^q \odot \text{pre}^d) + b_{s_2}\right). \tag{9}$$

Given a pair of weakly supervised training samples, we use the hinge loss [42] of information retrieval as the loss function, which is calculated as follows:

$$\begin{aligned}
\text{loss} = \sum_{(q, d^+, d^-) \in \{\text{Query}, D^+, D^-\}} & \max(0, 1 - (f_1(q, d^+) + f_2(q, d^+)) \\
& + (f_1(q, d^-) + f_2(q, d^-))),
\end{aligned} \tag{10}$$

where $f_1(q, d)$ represents the interactive matching score between the query and the document. $f_2(q, d)$ represents the semantic matching score between the query and the document to balance human-judged relevance labels and BM25 model scores. The model is fine-tuned using human-labeled relevance judgments after the parameters of the network is pretrained using the weakly supervised data.

4. Experiments Settings

This section describes our experimental dataset, metrics, baselines, and other implementation details.

4.1. Dataset. The experimental datasets come from Semantic Scholar (<http://corpus.semanticscholar.org/>) provided by Xiong et al., including 170,983 candidate document set (<https://alleninstitute.org/>) with 100 queries. Manual relevance judgement (a 5-level scale, including 4, 3, 2, 1, 0, 4 is the most relevant and 0 is not relevant) are available (<http://boston.lti.cs.cmu.edu/appendices/WWW2016/>) on the dataset, where both the relevant and irrelevant documents are labeled for each query. In this work, we mainly use the title and abstract of the paper as in [4]. The reasons are as follows: first, the title and abstract of the paper can refine the main points of the paper and summarize the main content of the original text. Second, the original text of the paper is not easy to obtain. Third, the original paper has a large amount of data and slow processing speed. The statistics of

experimental datasets is shown in Table 1. Ultimately, we convert this data to a unified format and use this data as an analog for real academic IoT search.

4.2. Baselines and Metrics. We compare information retrieval methods that are popular in the field of information retrieval, including K-NRM [26], Conv-KNRM [27], MatchPyramid [24], DRMM [25], MV-LSTM [43], and ArcII [44].

KNRM: It first computes cosine similarity between query word and document words using a translation layer, and then, it uses a feed-forward network to perform kernel pooling to compute the relevance score between query and document.

Conv-KNRM: Conv-KNRM improves KNRM by using CNN filters to model n -gram soft matches of queries and documents to capture n -gram such as phrases, concepts, or entities existed in the queries and documents.

MatchPyramid: It computes pair-wise dot product between query and document word vectors to compute an interaction matrix. It then passes this matrix through CNN layers with dynamic pooling to compute the similarity score.

DRMM: It firstly maps the local interaction matrix of query and document into a fixed length matching histogram. Then, it uses forward matching network to learn hierarchical matching features. These features are calculated by term gate network to get the global relevance score of query and document.

MV-LSTM: It uses the word embeddings obtained by passing the sentences through a Bi-LSTM and then computes an interaction vector using cosine similarity or a bilinear operation. It finally passes the interaction vector through a feed-forward network to compute the similarity score.

ArcII: ArcII first computes the interaction feature vector between query and document using CNN layers. It then computes the score for the query-document interaction vector using feed-forward network.

In academic search, the quality of top-ranked results is critical to improve user satisfaction, so the top rankings of Precision and NDCG are particularly important. In order to evaluate the efficiency of academic retrieval model, we report four standard evaluation indicators: MAP, MRR, Precision of top 5, 10, 15, 20 documents retrieved (P@5, P@10, P@15, P@20), NDCG of top 5, 10, 15, 20 documents retrieved (N@5, N@10, N@15, N@20).

4.3. Experiments Setting and Training Details

4.3.1. Knowledge-Enhanced Representation. First, we employ SCIE to extract the entities and relationships between entities contained in each scientific document. A total of 12,326,751 triples are formed. The detailed knowledge extraction results are shown in Figure 4. From the figure, we can see the proportion of the extracted relationships between entities. We use word2vec based on skip-gram and TransE to train the knowledge-enhanced feature representation, which is optimized by negative sampling. We set the embedding dimension to be 300, and the window size to be 5.

4.3.2. Weakly Supervised Training Data Preparation. First, for each query in the given dataset, the manually labeled documents are formed into positive and negative sample

TABLE 1: The statistics of academic datasets.

Details	Number
Number of queries	100
Number of papers	170,983
Sum of title length	1,200,641
Sum of abstract length	43,241,159

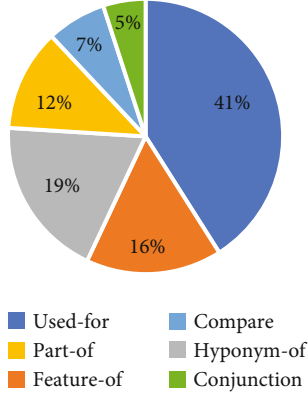


FIGURE 4: Proportion of relationship between entities in the structured knowledge.

pairs according to the label level. Then, we use the BM25 algorithm to recall 500 documents with higher scores from the candidate documents. In these recalled 500 documents, the documents which have been manually labeled are marked as positive samples, and other documents are marked as negative samples to form positive and negative sample pairs. Then, we use the title of a paper in the document collection as a pseudoquery, and its corresponding document as a positive example. Then, we use BM25 algorithm to search for the corresponding negative examples of this pseudoquery. Finally, a weakly supervised training set is formed, including original labeled data (query, positive samples, negative samples) and pre-labeled data (prequery, prepositive samples, and prenegative samples). The model is pretrained on weakly supervised training data and fine-tuned on human-labeled training data.

4.3.3. Effectiveness Verification Settings of the Proposed Model.

Our model is abbreviated as KER-IPM-WS. It includes three modules named as knowledge-enhanced representation learning module (KER), matching module (ISM, including interactive matching IM and semantic matching SM), and weakly supervised training (WS). In order to verify the effectiveness of each module, we designed six model variants as follows:

KER-IM: Academic search is implemented based on knowledge-enhanced feature representation, the academic text matching of which is based on interactive matching.

KER-SM: Academic search is implemented based on knowledge-enhanced feature representation; the academic text matching of which is based on interactive matching and semantic matching.

KER-ISM: Academic search is implemented based on knowledge-enhanced feature representation; the academic text matching of which is based on semantic matching.

KER-IM-WS: Weakly supervised academic search is implemented based on knowledge-enhanced feature representation; the academic text matching of which is based on interactive matching.

KER-SM-WS: Weakly supervised academic search is implemented based on knowledge-enhanced feature representation; the academic text matching of which is based on semantic matching.

KER-ISM-WS: Weakly supervised academic search is implemented based on knowledge-enhanced feature representation; the academic text matching of which is based on interactive matching and semantic matching.

We implement our model using TensorFlow. For the neural retrieval models CDSSM, KNRM, and Conv-KNRM, we use an open-source implementation MatchZoo (<https://github.com/NTMC-Community/MatchZoo>). For KNRM and Conv-KNRM, we set the number of bins to 11. We apply gradient descent algorithm and Adam as our optimizer for training the ranking model. And we use dropout technology to prevent overfitting. We set the batch size to 64 and select the learning rate from $[1e-1, 1e-2, 1e-3, 1e-4]$. The training epoch number is set to 20. We use 5-fold crossvalidation to validate our model. We randomly split all training data into five equal partitions. In each fold, three partitions are used for training, one for validation and one for testing.

5. Result Analysis and Use Case

5.1. Result Analysis. In this section, we first verify the effectiveness of knowledge-enhanced representation by comparing with the retrieval methods in baselines. Then, we further compare the academic retrieval performance of different module of KER-ISM-WS and analyze the impact on retrieval performance of different module of KER-ISM-WS.

5.2. Retrieval Performance of Different Retrieval Model.

Tables 2 and 3, respectively, show the retrieval performance of weakly supervised academic search based on knowledge-enhanced feature representation and the baselines on $NDCG@\{5, 10, 15, 20\}$ and $Precision@\{5, 10, 15, 20\}$. Table 4 shows the retrieval performance of the proposed weakly supervised academic search based on knowledge-enhanced feature representation and the baselines on MAP and MRR.

It can be seen from Tables 2, 3, and 5, the retrieval performance of our model exceeds the baselines. Experimental results show that our method is effective for academic search. It is worth noting that ConV-KNRM employs CNN filters to model n -gram of queries and documents to identify phrases, concepts, or entities. In addition to our model, ConV-KNRM achieves the best retrieval performance in the baselines. These findings indicate that mining phrases, concepts, or entities existed in academic texts is helpful for improving academic retrieval performance.

TABLE 2: Performance comparison on NDCG of different retrieval models.

Model	N@5	N@10	N@15	N@20
KNRM	0.4160	0.4807	0.5406	0.6077
ConvKNRM	0.4437	0.5258	0.5891	0.6535
MatchPyramid	0.3655	0.4475	0.5056	0.5727
DRMM	0.3935	0.4536	0.5411	0.5996
MV-LSTM	0.3244	0.4013	0.4822	0.5525
ArcII	0.4256	0.4768	0.5587	0.6094
KER-IPM-WS	0.5842	0.6009	0.6396	0.6653

TABLE 3: Performance comparison on precision of different retrieval models.

Model	P@5	P@10	P@15	P@20
KNRM	0.5241	0.5103	0.4987	0.4653
ConvKNRM	0.5452	0.5327	0.5048	0.4852
MatchPyramid	0.4517	0.4310	0.4321	0.3978
DRMM	0.4647	0.4414	0.4187	0.3831
MV-LSTM	0.4207	0.4034	0.3895	0.3527
ArcII	0.4655	0.4264	0.4207	0.3862
KER-IPM-WS	0.6509	0.6232	0.5521	0.5248

TABLE 4: Performance comparison on NDCG of different module of KER-ISM-WS.

Model	N@5	N@10	N@15	N@20
KER-IM	0.4753	0.5075	0.5672	0.6157
KER-SM	0.4427	0.4869	0.5406	0.5993
KER-ISM	0.5293	0.5693	0.5919	0.6269
KER-IM-WS	0.5228	0.5452	0.5816	0.6436
KER-SM-WS	0.4865	0.5133	0.5791	0.6255
KER-ISM-WS	0.5842	0.6009	0.6396	0.6653

TABLE 5: Performance comparison on MAP and MRR of different retrieval models.

Model	MAP	MRR
KNRM	0.4173	0.4929
ConvKNRM	0.4521	0.5402
MatchPyramid	0.3849	0.4730
DRMM	0.4158	0.4775
MV-LSTM	0.3437	0.4208
ArcII	0.4485	0.5086
KER-IPM-WS	0.5569	0.6427

5.3. *The Impact on Retrieval Performance of Different Module of KER-ISM-WS.* This section shows the impact of different module of KER-ISM-WS on retrieval performance Precision, NDCG, MAP, and MRR. The results are shown in Tables 4 and 6.

TABLE 6: Performance comparison on precision of different module of KER-ISM-WS.

Model	P@5	P@10	P@15	P@20
KER-IM	0.5733	0.5409	0.4961	0.4628
KER-SM	0.5439	0.5028	0.4874	0.4379
KER-ISM	0.5931	0.5217	0.5037	0.4867
KER-IM-WS	0.6031	0.5478	0.4856	0.4701
KER-SM-WS	0.5834	0.5254	0.4713	0.4443
KER-ISM-WS	0.6509	0.6232	0.5521	0.5248

TABLE 7: Performance comparison on MAP and MRR of different module of KER-ISM-WS.

Model	MAP	MRR
KER-IM	0.4764	0.5921
KER-SM	0.4491	0.5317
KER-ISM	0.5252	0.6437
KER-IM-WS	0.5121	0.5958
KER-SM-WS	0.4964	0.5838
KER-ISM-WS	0.5569	0.6427

5.3.1. *The Impact on Retrieval Performance of Knowledge-Enhanced Feature Representation.* KER-IM is academic search based on knowledge-enhanced feature representation; the academic text matching of which is based on KNRM. Therefore, we can compare it with KNRM to verify the effectiveness of the knowledge-enhanced representation. By comparing the retrieval performance of KNRM in Tables 2, 3, and 5 and KER-IM in Tables 4, 6, and 7, we can conclude that the proposed knowledge-enhanced feature representation method can improve retrieval performance.

5.3.2. *The Impact on Retrieval Performance of Matching Based on Interactive Matching and Semantic Matching.* We can compare KER-IM, KER-SM, and KER-ISM to verify the effectiveness of the matching based on interactive matching and semantic matching. From Tables 4, 6, and 7, the results indicate that the retrieval performance of the model KER-ISM is better than KER-IM and KER-SM. The reason is that KER-ISM considers both word interaction matching and sentence semantic matching to guide more accurate search.

5.3.3. *The Impact on Retrieval Performance of Weakly Supervised Training.* We can compare KER-IM and KER-IM-WS, KER-SM and KER-SM-WS, and KER-ISM and KER-ISM-WS to verify the effectiveness of the weakly supervised training. From Tables 5–7, the results indicate that the retrieval performance of the models KER-IM-WS, KER-SM-WS, and KER-ISM-WS is better than that of KER-IM, KER-SM, and KER-ISM. The possible reason is that a large amount of weakly supervised training data can guide the model to learn better parameters.

TABLE 8: Search results of query “object detection”.

KER-PM-WS	Paper title	Human label	Entity relation
1	Rapid object detection using a boosted cascade of simple features	4	Cascade of simple features USED-FOR object detection
2	Class-specific Hough forests for object detection	5	Hough forests USED-FOR object detection
3	Faster R-CNN: towards real-time object detection with region proposal networks	3	R-CNN USED-FOR object detection
4	Learning rich features from RGB-D images for object detection and segmentation	2	Learning rich features USED-FOR object detection RGB-D images USED-FOR learning rich features RGB-D images USED-FOR object detection
5	Rich feature hierarchies for accurate object detection and semantic segmentation	3	Rich feature hierarchies USED-FOR object detection

Query: object detection.

5.4. Use Case—Academic Search. In this section, we use the proposed KER_ISM_WS model to report the academic search results. A real use case for literature retrieval is as follows in Table 8. The input query is object detection, which reflects the information needs of users.

Table 8 shows the retrieval results of our model and the manually labeled level for these retrieval results. Entities in the titles of the retrieved papers are highlighted in green. Our method also gives the entities that exist in the title of the paper and the relationships between them. Entities are marked as green; relationships are marked as red.

From the search results and manually labeled tags, we can see that our model has achieved accurate matching. At the same time, we infer from the manually annotated data that the query is trying to find documents that have a certain relationship with the “object detection.” The search results show that our model can retrieve articles which meet the needs. Moreover, all the items in the search results can express the entities related to the query and the relationship information between the entities. This further illustrates the effectiveness of the proposed representation learning model. It not only models the entities but also models the semantic relationships that exist between entities, which helps to improve the performance of the retrieval model.

6. Conclusion

We design academic Internet of Things search framework based on Internet of Things technology. In order to alleviate the pressure of the cloud server processing massive academic big data, the edge server is introduced to clean and remove the redundancy of the data to form a clean data for further analysis and processing by the cloud server. For academic search, we propose a novel knowledge-enhanced feature representation learning method which can express rich semantics in texts of academic search field. Aiming at the “data hungry” property of deep neural academic retrieval methods, we propose a weakly supervised academic search model based on the knowledge-enhanced representation, which

relieves high cost of acquisition of manually labeled data by obtaining a lot of pseudolabeled data in the process of academic search. In the process of text matching for academic retrieval, the proposed model considers both the word-level interactive matching and the sentence-level semantic matching to improve matching accuracy of relevance. Experiment results on real academic search datasets show that the proposed model is effective and greatly improves the academic search performance.

Data Availability

The datasets we use in this paper can be downloaded from <http://boston.lti.cs.cmu.edu/appendices/WWW2016/>.

Conflicts of Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgments

This work was supported by the National Key R&D Program of China (2018YFB1402600), National Natural Science Foundation of China (NSFC) under grant (No. 61772083, No. 61802028, and No. 61877006).

References

- [1] J. Zhang and J. Tang, “Name disambiguation in AMiner,” *Science China-information sciences*, vol. 64, no. 4, 2021.
- [2] G. Abramo, C. A. D’Angelo, and F. di Costa, “A gender analysis of top scientists’ collaboration behavior: evidence from Italy,” *Scientometrics*, vol. 120, no. 2, pp. 405–418, 2019.
- [3] X. Chenyan, P. Russell, and C. Jamie, “Explicit semantic ranking for academic search via knowledge graph embedding,” in *Proceedings of the 26th International Conference on World Wide Web*, Perth, Australia, 2017.

- [4] J. Shen, X. Jinfeng, X. He, S. Jingbo, S. Saurabh, and H. Jiawei, "Entity set search of scientific literature: an unsupervised ranking approach," in *Proceedings of the 41th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Ann Arbor, MI, USA, 2018.
- [5] S. Marchesin, A. Purpura, and G. Silvello, "Focal elements of neural information retrieval models. An outlook through a reproducibility study," *Information Processing and Management*, vol. 57, no. 6, article 102109, 2020.
- [6] K. Zhang, C. Xiong, Z. Liu, and Z. Liu, "Selective weak supervision for neural information retrieval," in *Proceedings of The Web Conference*, Taipei, Taiwan, 2020.
- [7] L. Luan, L. He, O. Mari, and H. Hannaneh, "Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium, 2018.
- [8] S. Liang and C. Y. Huang, "GeoCENS: a geospatial cyberinfrastructure for the world-wide sensor web," *Sensors*, vol. 13, no. 10, pp. 13402–13424, 2013.
- [9] S. Pattar, R. Buyya, K. R. Venugopal, S. S. Iyengar, and L. M. Patnaik, "Searching for the IoT resources: fundamentals, requirements, comprehensive review, and future directions," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2101–2132, 2018.
- [10] L. Chen, D. Chen, F. Yang, and J. Sun, "A deep multi-task representation learning method for time series classification and retrieval," *Information Sciences*, vol. 555, pp. 17–32, 2021.
- [11] M. Tomas, S. Ilya, C. Kai, C. Greg, and D. Jeffrey, "Distributed representations of words and phrases and their compositionality," in *Proceedings of the 27th Annual Conference on Neural Information Processing Systems*, Lake Tahoe, Nevada, USA, 2013.
- [12] P. Jeffrey, S. Richard, and D. Manning Christopher, "Glove: global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, 2014.
- [13] M. Bhaskar and C. Nick, "An introduction to neural information retrieval," *Foundations and Trends in Information Retrieval*, vol. 13, no. 1, pp. 1–126, 2019.
- [14] D. Hongliang, T. Siliang, and F. Wu, "Entity mention aware document representation," *Information Science*, vol. 430, pp. 216–227, 2018.
- [15] O. P. Ghasnezhad, K. Wang, and Z. Wang, "An embedding-based approach to rule learning in knowledge graphs," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 4, pp. 1348–1359, 2021.
- [16] J. Huang, H. Wang, W. Zhang, and T. Liu, "Multi-task learning for entity recommendation and document ranking in web search," *ACM Transactions on Intelligent Systems and Technology*, vol. 11, no. 5, pp. 1–24, 2020.
- [17] T. Komamizu, "Random walk-based entity representation learning and re-ranking for entity search," *Knowledge and Information Systems*, vol. 62, no. 8, pp. 2989–3013, 2020.
- [18] C. Xiong, Z. Liu, J. Callan, and T. Liu, "Towards better text understanding and retrieval through kernel entity salience modeling," in *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, Ann Arbor, MI, USA, 2018.
- [19] C. Xiong, J. Callan, and T. Liu, "Bag-of-entities representation for ranking," in *Proceedings of the 2016 ACM International Conference on the Theory of Information Retrieval*, Newark, DE, USA, 2016.
- [20] H. Raviv, O. Kurland, and D. Carmel, "Document retrieval using entity-based language models," in *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Pisa, Italy, 2016.
- [21] C. Xiong, J. Callan, and T. Liu, "Word-entity duet representations for document ranking," in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Shinjuku, Tokyo, Japan, 2017.
- [22] Z. Liu, C. Xiong, M. Sun, and Z. Liu, "Entity-duet neural ranking: understanding the role of knowledge graph semantics in neural information retrieval," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Melbourne, Australia, 2018.
- [23] Y. Shen, X. He, J. Gao, L. Deng, and G. Mesnil, "Learning semantic representations using convolutional neural networks for web search," in *Proceedings of the 23th International Conference on World Wide Web*, Seoul, Korea, 2014.
- [24] P. Liang, Y. Lan, J. Guo, J. Xu, S. Wan, and X. Cheng, "Text matching as image recognition," in *Proceedings of the 30th AAAI Conference on Artificial Intelligence*, Phoenix, Arizona, USA, 2016.
- [25] J. Guo, Y. Fan, Q. Ai, and W. Croft, "A deep relevance matching model for ad-hoc retrieval," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, Indianapolis, IN, USA, 2016.
- [26] C. Xiong, Z. Dai, J. Callan, Z. Liu, and R. Power, "End-to-end neural ad-hoc ranking with kernel pooling," in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Shinjuku, Tokyo, Japan, 2017.
- [27] Z. Dai, C. Xiong, J. Callan, and Z. Liu, "Convolutional neural networks for soft-matching n-grams in ad-hoc search," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, Marina Del Rey, CA, USA, 2018.
- [28] Y. Zheng, Y. Liu, Z. Fan et al., "Investigating weak supervision in deep ranking," *Data and Information Management*, vol. 3, no. 3, pp. 155–164, 2019.
- [29] H. Zamani and W. Croft, "On the theory of weak supervision for information retrieval," in *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval*, Ann Arbor, MI, USA, 2018.
- [30] R. Levy, B. Bogin, S. Gretz, R. Aharonov, and N. Slonim, "Towards an argumentative content search engine using weak supervision," in *Proceedings of the 27th International Conference on Computational Linguistics*, Santa Fe, New Mexico, USA, 2018.
- [31] J. Han, Y. Yang, D. Zhang, D. Huang, D. Xu, and F. de la Torre, "Weakly-supervised learning of category-specific 3D object shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 4, pp. 1423–1437, 2021.
- [32] H. Zamani and W. Croft, "Towards theoretical understanding of weak supervision for information retrieval," in *Proceedings of the ACM SIGIR Workshop on Learning from Limited or Noisy Data for Information Retrieval*, Ann Arbor, MI, USA, 2018.
- [33] B. Li, P. Cheng, and L. Jia, "Joint learning from labeled and unlabeled data for information retrieval," in *Proceedings of the 27th International Conference on Computational Linguistics*, Santa Fe, New Mexico, USA, 2018.

- [34] X. Ling, W. Deng, C. Gu, H. Zhou, C. Li, and F. Sun, "Model ensemble for click prediction in bing search ads," in *Proceedings of the 26th International Conference on World Wide Web Companion - WWW '17 Companion*, Perth, Australia, 2017.
- [35] L. Xue, Z. Jianjin, L. Zhipeng et al., "Learning fast matching models from weak annotations," in *Proceedings of the World Wide Web Conference*, San Francisco, CA, USA, 2019.
- [36] M. Dehghani, H. Zamani, A. Severyn, J. Kamps, and W. Croft, "Neural ranking models with weak supervision," in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Shinjuku, Tokyo, Japan, 2017.
- [37] S. Mac Avaney, K. Hui, and A. Yates, "An approach for weakly-supervised deep information retrieval," 2017, <http://arxiv.org/abs/1707.00189>.
- [38] S. Mac Avaney, A. Yates, K. Hui, and O. Frieder, "Content-based weak supervision for ad-hoc re-ranking," in *Proceedings of the 42th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Paris, France, 2019.
- [39] X. Sui, M. Li, Y. Ying et al., "Aerolysin nanopore identification of single nucleotides using the AdaBoost model," *Journal of Analysis and Testing*, vol. 3, no. 2, pp. 134–139, 2019.
- [40] T. H. Fereja, F. Du, C. Wang, D. Snizhko, Y. Guan, and G. Xu, "Electrochemiluminescence imaging techniques for analysis and visualizing," *Journal of Analysis and Testing*, vol. 4, no. 2, pp. 76–91, 2020.
- [41] A. Bordes, N. Usunier, A. García-Durán, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," in *Proceedings of the 27th Annual Conference on Neural Information Processing Systems*, Lake Tahoe, Nevada, USA, 2013.
- [42] J. Guo, Y. Fan, L. Pang et al., "A deep look into neural ranking models for information retrieval," *Information Processing and Management*, vol. 57, no. 6, p. 102067, 2020.
- [43] S. Wan, Y. Lan, J. Guo, J. Xu, P. Liang, and X. Cheng, "A deep architecture for semantic matching with multiple positional sentence representations," in *Proceedings of the 30th AAAI Conference on Artificial Intelligence*, Phoenix, Arizona, USA, 2016.
- [44] B. Hu, Z. Lu, H. Li, and Q. Chen, "Convolutional neural network architectures for matching natural language sentences," in *Proceedings of the 28th Annual Conference on Neural Information Processing Systems*, Montreal, Quebec, Canada, 2014.

Research Article

Innovation and Development of Rural Leisure Tourism Industry Using Mobile Cloud IoT Computing

Guangwei Wang 

Aviation and Tourism College, Guilin University of Aerospace Technology, Guangxi, China

Correspondence should be addressed to Guangwei Wang; wangguangwei198809@guat.edu.cn

Received 6 May 2021; Revised 26 July 2021; Accepted 11 August 2021; Published 26 August 2021

Academic Editor: Michel Kadoch

Copyright © 2021 Guangwei 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.

As a promising IoT application, the rural leisure tourism industry can promote the reconstruction of industrial structure in rural areas and realize a sustainable, rapid, and healthy development of rural economy. This paper takes the rural leisure tourism industry in China as an example and aims at building an intelligent and integrated modern IoT use case. Based on the traditional rural leisure tourism, we improve the system by adding the data analysis over a mobile cloud IoT computing platform. In particular, this work investigates the characteristics of the national tourism market under the security requirements from governmental cloud data management policy. Our study shows that the geographical concentration index G of tourists in the Chinese market continues to increase. With the booming of IoT applications in rural leisure tourism, intelligent and integrated tourism guidance and optimized decision-making will provide tourists with better information and thus make rapid improvement of geographical concentration index.

1. Introduction

Modern rural leisure tourism is a new type of industry that uses agricultural production and rural ecological processes to provide consumers with entertainment, sightseeing, and experience services. Historical data shows that it has also promoted the transformation and upgrading of the agricultural industry. The Internet of Things (IoT) is a key supporting technology for rural leisure tourism. For developed countries, leisure agriculture and rural tourism are the consequences of urban industrial pollution and fast-paced lifestyles. As early as the 1830s, rural agricultural tourism began in Europe. With the continuous spread of global epidemics in the past two years, the rural leisure tourism has shown new vitality and become one of the important directions of the global tourism industry.

Information technology-based tourism has been widely studied in the past. By using the anonymous mobile location information of national mobile network operator of Estonia, Chen conducted an empirical analysis on the field visit data of foreign tourists based on the dimensions of geography, time, and composition [1]. Dai, instead, utilized big data technology over detailed telephone record (CDR) data and

real-time location information of mobile phones, in order to monitor the flow of tourists in the scenic spot and analyze the tourist behavior in the scenic spot [2]. Based on big data, Zhou analyzed the influencing variables of the branding of tourism destinations on Facebook [3]. Qi used social media data from Flickr and Twitter to assess the potential tourism security threats to bird and biodiversity areas that are globally well-known and believes that the development of tourism can promote the protection of biodiversity areas [4]. Zhou conducted social network and semantic big data analysis on tourist content retrieved from online communities in the TripAdvisor forums in 7 major European capital cities and conducted a tourism demand forecast analysis [5].

In the meantime, Porrambage made use of Barcelona as an example to analyze the influence of social networks on Chinese tourists' travel behavior [6]. Smith collected real-time big data of Wi-Fi routers installed in 149 different locations on Jeju Island, South Korea, and analyzed the behavior patterns of tourists from 2016 to 2017. Through statistical analysis of the behavior patterns of tourists, he proposed an optimal Tour route recommendation system [7]. Charles-Edwards employed big data together with geographic location information in

Flickr social media to analyze the tourist attraction of six Italian cities [8]. Based on the GPS trajectory data of the social photo-sharing website (Flickr) and the network platform (Wikiloc), Sullivan took the Teide National Park as the research area, revealed the travel network and spatial distribution characteristics of tourists, and analyzed the social network geotag data [9].

The research mentioned above has made a significant contribution to the development of the tourism industry, but there is no in-depth study on mobile cloud IoT computing so far, and there is also a problem of insufficient awareness of the current cultural industry. China is a largely agricultural country with a long history, with a vast agricultural area, a profound agricultural culture, and strong rural customs. The development of rural leisure tourism has excellent conditions, strong demand, and broad prospects. In recent years, a lot of research has been conducted on the rural leisure tourism industry, but the research content mainly focuses on the competitiveness of the tourism industry, temporal and spatial distribution, market segmentation, influencing factors, and demand forecasting. Moreover, the existing research objects are relatively single, mainly limited to specific tourist destinations or tourist sources, and lack comprehensive systematic research.

The development of modern rural leisure tourism can provoke the readjustment of rural industrial structures. This is a powerful measure to achieve sustained, rapid, and healthy development of the rural economy. This is also of great significance for the sustainable development of society. However, rural leisure tourism data has not yet been effectively collected and processed. The existing limited tourism data is held by different agencies, such as tourism authorities, tourism data agencies, and cloud service providers. There is a lack of effective travel data exchange and sharing between these institutions. Therefore, various tourism data in different tourist areas should be safely collected, transmitted, and intelligently processed for overall decision-making and optimization analysis. All relevant government departments, agencies, and service providers must actively collaborate to meet the needs of cross-departmental data sharing in rural leisure tourism and ensure the security of user data and information. At the same time, rural leisure tourism should be constructed as a whole with the development of the modern rural economy, not just applied separately. If the modern rural economy is regarded as an ecosystem, rural leisure tourism should become an important link in this ecosystem.

This paper endeavors to build an intelligent and comprehensive application of the IoT for modern rural leisure tourism. A large number of IoT devices (such as various sensors, cameras, and cell phones) are used to monitor and collect various travel data and, then, aggregate them to the mobile-cloud IoT computing platform. We use cloud computing and machine learning to intelligently analyze and optimize tourism data and obtain the best decision-making results overall. We take the Chinese rural leisure tourism industry as the analysis object, introduce the tourism market based on cloud computing, and analyze foreign tourism data to enhance the diversity of rural culture. This work also highlights the security requirements of government cloud data management and studies the measures to achieve cloud tourism data security.

The rest of the paper is organized as follows. Section II presents the mobile cloud IoT computing platform for rural leisure tourism. Section III develops the machine learning algorithm to mine the tourism data collected. Section IV uses China's rural leisure tourism sector as an example to examine the features of the national tourist market, followed by Section V to conclude the paper.

2. Rural Leisure Tourism on Mobile Cloud IoT Computing Platform

Mobile Cloud IoT Computing has great limitations in processing rural leisure tourism industry data. Although academia has done some exploration, the problems faced by the practice department are still difficult to solve. The reason is that the existing research strategies are relatively scattered, and the tourism data cloud security is not systematically constructed as an organic whole, and the interaction between the elements of the security system is not explored relationship [10]. Most of the existing strategies start from a single perspective of tourism data institutions or cloud service providers and ignore the coordination mechanism among different entities such as tourism data institutions, cloud service providers, and government cloud management units. Existing studies generally discuss "cloud", and usually point to the public cloud environment and fail to make a detailed analysis of the government cloud environment nor distinguish the different scenarios of tourism data institutions using government cloud services [11]. The core concept of the rural leisure tourism industry is conducive to understanding the symbiotic evolution relationship between different subjects in the cloud of tourism data and the external environment such as institutional environment, institutional environment, and market environment. It has reference value for proposing a more systematic security strategy of the tourism data cloud. Therefore, this paper introduces the perspective of the ecosystem to explore the security of tourism data systems [12].

We take the national cloud computing guidance tourism market as the research subject. For the availability of data, in the specific analysis of the characteristics of the national tourist market, this paper carries out a semistructured research on the heads of 11 tourism data institutions in Beijing, Shandong, Jiangsu, Henan, Inner Mongolia, and other regions, so as to understand the construction status and security needs of the tourism data management system in the government cloud.

On the basis of the traditional rural leisure tourism industry, the cultural mode of the original tourism area is retained, and the data analysis and integration guidance of Mobile Cloud IoT Computing Internet of things are added to guide tourists. The data integration model is dominated by tourism data institutions, and the data is transmitted to tourists by the server through cloud computing.

As shown in Figure 1, the first scenario is that the tourism data organization (including the tourism data center and the agency tourism data room) migrates the completed tourism data management system from the local to the government cloud or builds a new tourism data management system based on the government cloud (the system is only for the internal use of the tourism data organization). The

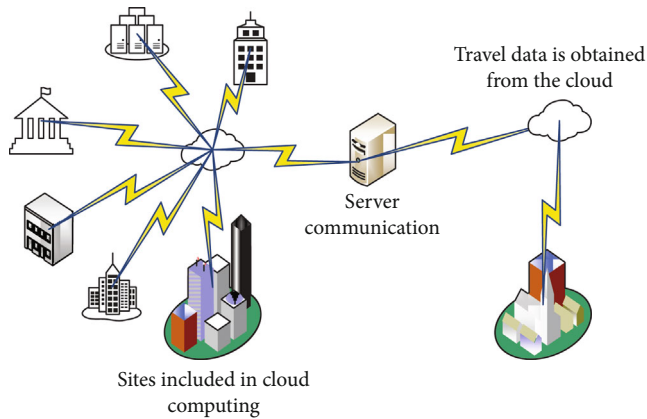


FIGURE 1: Cloud computing provides urban tourism data computing services.

second scenario is the construction of a unified tourism data management system based on government cloud by a regional comprehensive tourism data library under the organization and coordination of the tourism data authorities at the same level, including the tourism data management system for filing units and the digital tourism data library system for tourism data library, which provides the same level authorities' tourism data room and other comprehensive tourism data in the region. In different scenarios, producers, consumers, and decomposers refer to slightly different subjects. Based on the ecological model of tourism data cloud security at the macro-level, this paper puts forward the security strategy for the tourism data security in the government cloud environment.

Based on the public cloud analysis model, this paper constructs the configuration model of the public cloud planning layout, which promotes the optimal allocation of public cloud resources under the objectives and constraints of service level agreement. Three types of resource pools are deployed according to different requirements of business scenarios: low-cost centralized deployment mode is adopted for cold scenic spot data and warm scenic spot data business to reduce cost and improve efficiency; in hot scenic spot data scenario, regional centralized deployment mode is adopted for high-density business and business side deployment mode is adopted for low-density business to ensure service quality according to the different business density of each region. At first, the public cloud is mainly used in telecom services, and the purpose is to create a healthy network ecological environment, so that tourists can enjoy better services and protect the rights and interests of tourists. Facing the practical problems in the planning of public cloud, public cloud analysis provides the idea of transmitting the business side service commitment to the back-end resource allocation ability, so that the business and resources can be configured simultaneously and effectively support the development of public cloud business.

2.1. Demand Division of Cloud Computing in Tourism Planning Calculus. Passenger flow is primarily focused in summer and autumn due to apparent seasonal variations in regional passenger flow, and there is an inadequate supply of tourist goods in spring and winter. By introducing four-season tourist goods,

the tourism management department should take the appropriate steps to minimize the difference in tourism seasons and scientifically divert, steer, and control passenger flow. First and foremost, we will organize a regional ice and snow festival with the theme "Ice and Snow Outside the Great Wall, Silk Road Wonderland," and promote desert ice and snow tourism as a new winter and tourist business card. The size and appeal of ice and snow tourist goods are enhanced by spring tourism [13]. Northern China has released a succession of preferential policies, preferential policies, and subsidized policies for winter and spring tourism goods in recent years, but they have not boosted the market's vitality and are not favorable to tourist diversion flow. Diverse cultural tourist events will be organized at various peak periods in the future to bridge the gap between low and high seasons [14]. Increase participatory tourism experience activities incorporating historical and cultural elements, expand the play space and stay time of tourists in the region, and promote the West by creating RV campgrounds, customs outside the customs, or desert theme hotels, scientifically arrange certain tourist reception service facilities, increase participatory tourism experience activities incorporating historical and cultural elements, expand the play space and stay time of tourists in the region, and promote the West by creating RV campgrounds, customs outside the customs, or desert theme hotels. The route changes from tourist goods to participation experience tourism products, strengthening western route tourism products and decreasing the regional tourism flow's "core-periphery" hierarchical structure [15]. Taking the construction of the National Tourism Demonstration Zone as an opportunity, we will realize the efficient reorganization of tourism products and routes on the east-west line and further strengthen the construction of the transportation network to form the overall effect of tourism products. Take the western line passenger flow as an example, use the new media big data platform to promote the effective management of the tourist destination space and enhance the travel accessibility and radiation effect [16]. To promote effective management of tourist administration departments, use big data, new media, and smart tourism as platforms, it is feasible to disclose the visitor flow trajectory and tourist behavior preferences with high accuracy using tourist data information, to scientifically design tourist routes, to encourage efficient and precise destination marketing, and to further increase the potential tourist market. Promote the safe and healthy growth of the tourism sector in order to minimize the strain on visitor flow, particularly during the peak season of World Heritage tourism.

For cloud service providers, centralized deployment and decentralized deployment are indispensable. Telecom Tianyi cloud adopts the typical "centralized + decentralized" mode, but it is difficult to determine the matching principle of the two types of resource pools in the actual planning. Due to the utility characteristics of cloud, cloud tourists and cloud resource environment will constantly change. How to plan, schedule, and adjust cloud resources in a dynamic environment to meet the complex needs of tourists has become an urgent problem in a cloud computing environment. Therefore, cloud resource layout planning and design must evaluate business needs, predict business development trends and adjust the layout, in order to balance operating costs and

customer perception, and occupy a favorable position in the competition. The significance of service level agreement (SLA) between cloud tourists and cloud service providers is not only to protect the rights and interests of customers but also to reveal the needs of tourists to service providers. How to transmit the business side service commitment to the back-end resource allocation strategy and promote the optimal allocation of cloud resources under the objectives and constraints of SLA is the key to solve the problem of cloud resource allocation. It is an effective way to solve the problem of the source distribution. Based on the different needs of tourists, business scenarios are different resources are configured to maximize the satisfaction of tourists and resource utilization of cloud service providers. According to the different needs of tourists, public cloud business scenarios are divided into three categories: cold, hot, and warm. For typical industries and key products, cloud business scenarios are classified, and different resource requirements (delay, performance, etc.) of SLA levels are formulated.

3. Machine Learning-Based Tourism Data Analysis

3.1. Tourism Data Security Based on Machine Learning. To meeting the requirement of tourism data protection, this paper explores the security cooperation mechanism between tourism data authorities and tourism data preservation institutions. Since 2018, most of the local tourism bureaus have been actively reformed to set up their own data centers. Some interviewees said that after the reform, the competent departments of tourism data perform the administrative functions and do not assume the responsibility of preserving tourism data [17]. There is no strong demand for tourism data management based on the government cloud. After the institutional reform, the competent departments of tourism data are generally short of manpower, but they have to undertake the functions of four or five departments, including professional title evaluation, education and training, law enforcement, and inspection [18]. In addition, most of the staff do not have a professional background in tourism data science or working experience in tourism data, so they have a negative attitude towards the security risks and avoidance strategies of tourism data in the government cloud environment [19]. In this case, the competent department of tourism data should strengthen the cooperation and contact with the tourism data center and the agency tourism data room, reach an agreement on the security issues of the application of government cloud by tourism data institutions, and strengthen the cooperation [20].

We will improve the legal system, standards, and norms and promote pilot projects. Based on the opinions, on the one hand, the competent departments of tourism data should timely issue operational system specifications to guide the risk identification and assessment of tourism data institutions and provide reference framework and methodological support for the formulation of risk response strategies; on the other hand, according to the superposition of security needs in the political cloud environment, they should fully implement the system specifications of tourism

data security considering the security risks brought by the government cloud, and this paper proposes and refines the tourism data security requirements applicable to the government cloud environment [21]. The concept of superposition of security requirements is not a unique product in the cloud environment, but the traditional informatization [22].

Through application review, expert review, and third review, the tourist data authority can advise the comprehensive tourism data center and government tourism data room on how to choose system developers and cloud service agents. Improve the quality of service provided by tourist data information technology and content management software businesses, as well as the provision of tourism data information-related systems, platforms, tools, and other goods. Many respondents indicated that localized goods are not cost-effective, especially in terms of localized replacement, and that there is a conflict between safety and cost-effectiveness. Based on attaining safety, dependability, and self-control, relevant firms should enhance the functional integrity and performance stability of their goods [23].

3.2. Interactive Mechanism Model of Tourism Data. Aiming at the innovation and development of the tourism rural leisure tourism industry, this paper constructs the interactive mechanism model of tourism data subject combined with cloud computing [24]. Tourism data institutions, cloud service providers, and government cloud management units are all important components of tourism data cloud security ecology, and the construction of subject interaction mechanism needs Mobile Cloud IoT Computing [25]. Among them, the Internet is the key, focusing on clear security responsibilities and responsibility boundaries; cloud computing is the foundation, focusing on communication and cooperation among the three; supervision and audit is an important way to build a new relationship among them, and let n be the number of boundary points of cloud computing, then the relationship strength G can be expressed as

$$G = \frac{\sum_{Z=1}^{h_j} \sum_{t=1}^{n_h} |y_{ji} - y_{hr}|}{n_j n_h (u_j + u_h)}, \quad (1)$$

$$l = 1 - \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}. \quad (2)$$

Among them, n represents the annual increase in the number of visitors in the society, u is the number of visitors affected by the Internet, Z is the counting unit, and h is the peak value of visitors. In the formula (2), the subscript of u is the visitor integral interval, and C_1 and C_2 are the scattered point identifications of tourists in the integral interval, y is the end of the error, n is the number of nodes, and μ is the error representation of the number of visitors. In formula (2), σ is a negative error coefficient, and x and y are space point elements, respectively, and l is the degree of spatial relationship. This paper mainly uses the theoretical perspective of the ecosystem to carry out research and applies it at the macro and micro levels. At the macrolevel, we use the

natural ecosystem theory in the field of ecology for reference to build a tourism data security model. The following equations calculate the core index (φ) used to reflect the spatial concentration of cloud computing guiding the tourism market:

$$\varphi_k = \frac{2k}{k+1} + \left[\frac{1}{2} + \frac{1}{2k} \right] \left[\frac{c_2 - c_1}{3} \right]^2 + \frac{2(c_2 - c_1)}{3}, \quad (3)$$

$$D_{jh} = \int_0^{\infty} dF_j(y) \int_0^y (y-x) dF_h(x), \quad (4)$$

$$G_t = \sum_{j=2}^k \sum_{h=1}^{j-1} G_{jh} (p_j s_h + p_h s_j) D_{jh} (1 - D_{jh}), \quad (5)$$

where G is the number of tourists from the k country of origin, the meaning of C in the formula is the same as formula (2), t is the total number of inbound tourists from China in that year, and N is the number of countries of origin. $F()$ is the tourist error function, and d is the representation of the integral, where it represents the integral of $F()$. The above formula (5) expresses the relationship between G and the integration result “ D_{jh} ” of formula (4), j , x , and y are the identifiers of the elements. The meaning of G here is the same as formula (1). The greater the G value, the more concentrated the spatial distribution of tourists, and the greater China’s dependence on the market. It can be used to analyze the competition level and development potential of the major source countries in the cloud computing-guided tourism market. We further make the following definitions:

$$\begin{aligned} f(x) &= \frac{1}{Nh} \sum_{i=1}^N k \left(\frac{X_i - x}{h} \right), \\ k(x) &= \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{x^2}{2} \right), \end{aligned} \quad (6)$$

where $f(x)$ and Nh represent the occupancy rate and the growth rate, $X_i - x$ is the first time that year h number of tourists from each country of origin, X_i is the number of tourists in the source country last year, and N is the total number of Chinese Inbound Tourists in that year. The average values of α and β in the period are calculated as $k(x)$, and the two-dimensional coordinates are divided to obtain the competitive state of each customer market in the country.

In this work, we utilize a specific time series model, a method to predict the future by using past observations, which can effectively predict the development trend of national cloud computing to guide the tourism market. The general form is as follows:

$$\ln \left(\frac{x}{x-1} \right) = \beta \ln x - 1, \quad (7)$$

$$\psi = \sum_{x=1}^{\theta} Vx = \sum_{x=1}^{\theta} \left(\frac{Wx}{\sum_{x=1}^{\theta} Wx} Sx \right). \quad (8)$$

In formula (7), x is the integral function and β is the time coefficient. In formula (8), the meaning of x is the same as formula (7), W is the difficulty coefficient of conventional guidance, and θ is the service area of the tour guide. Grey correlation analysis is a method to evaluate the degree of correlation between various indicators, which can be used to analyze the impact of various factors on the national tourists’ cloud computing guidance tourism demand. The steps are as follows: first, determine the reference sequence h_t and comparison sequence $w_c x_t$. The relationship between them is as follows:

$$h_t = \tan h(w_c x_t + u_c(r_t \Theta h_{t-1}) + b_c), \quad (9)$$

$$h_t = z_t \Theta h_{t-1} + (1 - z_t) \Theta h_t. \quad (10)$$

In formula (10), z is the coefficient of determination, corresponding to h , t is the increment of time, and Θ is the direction conversion of the vector to maintain the validity of the result h_t .

Second, the data are dimensionless, in formula (11), L represents the complex coefficient of curve processing, and p and q , respectively, represent the value of the trigonometric function mapping relationship in the prediction.

$$\theta(p, q) = \arctan \left(\frac{L(p, q+1) - L(p, q-1)}{L(p+1, q) - L(p-1, q)} \right). \quad (11)$$

Finally, the correlation coefficient and correlation degree are calculated in the following:

$$\begin{aligned} \ln \left(\frac{FI_{it}}{FI_{it} - 1} \right) &= \alpha + \beta \ln FI_{it} - 1, \\ k_{t1}[i] &= \sum_j \cos(w_i^1, w_j^2). \end{aligned} \quad (12)$$

Among them, α is the coefficient of rural tourism in the previous year, β is the coefficient predicted this year, and i and j indicate the difficulty coefficient of unlimited tempering. The degree of grey correlation FI represents the influence of cloud computing guided tour demand, which has a positive correlation.

4. Numerical Results and Analysis

4.1. Tourism Data Analysis with Mobile Cloud IoT Computing. Using the formula of entropy method to calculate the rural tourism development potential and its subsystem index of county unit in the region, the calculation results are shown in Table 1.

The measurement results are shown in Figure 2. The regional rural tourism development potential index is the highest (0.572), which is about 5 times of the lowest. Further statistics show that the overall development potential of rural tourism in the region is relatively low and the difference is obvious, among which 59.09% of the counties and districts have lower development potential than the average. Comparing the distribution of the development index of the

TABLE 1: Rural Tourism Development Potential and Its Subsystem Index.

Item	NSSP	Strategy	Government affairs	Cloud environment	Legal system
Inbound	1.15	1.83	0.52	0.62	0.76
Visitors	3.97	1.21	3.49	2.26	2.03
Geography	3.85	2.03	3.6	4.37	5.47
Concentration	2.83	2.66	4.83	4.44	5.9
Specification	1.49	3.08	2.97	3.61	2.92
Project pilot	2.82	6.56	4.17	4.09	5.45
Market	5.41	2.76	1.11	1.51	1.42

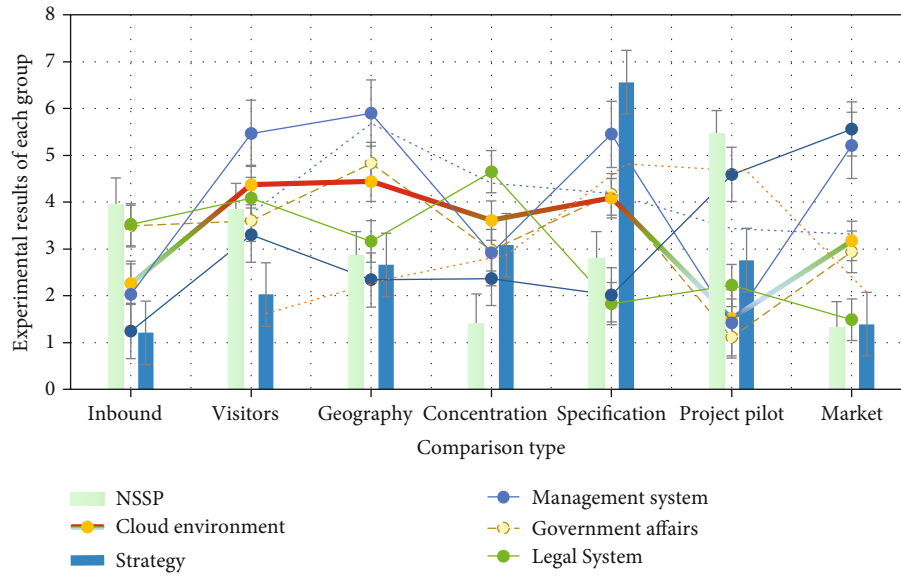


FIGURE 2: Regional rural tourism development potential index.

five subsystems, we can see that the conditional coefficient of variation of rural tourism elements is the largest, followed by the economic support of rural tourism and the tourism industry foundation of rural tourism region, while the endogenous development level of rural tourism is relatively small, and the variation coefficient of rural tourism environment background and rural tourism development potential is relatively close.

As shown in Figure 3, about 5.6% of the high-level rural tourism development potential areas are in the district and county units. Most of these areas are located in cities and autonomous prefectures divided into districts, which have a good tourist market environment and regional economic development foundation. The development potential index of rural tourism is more than 0.5. It is found that the high development potential of rural tourism is mainly due to the outstanding performance of A-level scenic spots, the number of traditional villages, and the number of star-rated farmhouse and star-rated rural tourism inns. The improvement of rural tourism reception capacity and supporting service facilities in this grade area provides a good foundation for the development of rural tourism.

As shown in Figure 4, the industrial base of rural tourism, the number of regional tourists and tourism economic

income, high star hotels, and other indicators also show obvious advantages, providing strong competitiveness for the development of rural tourism. Further exploring the elements of each subsystem, it is found that the background index of rural tourism environment in high-level development potential areas is generally low, which also restricts the further improvement of rural tourism development potential in this area to a certain extent.

The security of tourism data in e-government cloud environment is a complex system engineering, involving many subjects and elements. However, most of the existing literature tends to put forward suggestions and requirements for tourism data institutions and cloud service providers, respectively, ignoring the relevance. Therefore, based on the theoretical framework of the natural ecosystem and the results of cloud computing coding, this paper constructs a macrolevel ecological model of tourism data cloud security. The model focuses on the interaction between the subject and the environment and the interaction between the subjects. The security environment includes a management system, legal system, standards, project pilot, market guidance, and other specific environments. The security subject refers to the organization that undertakes the responsibility of tourism data security and can understand the tourism data

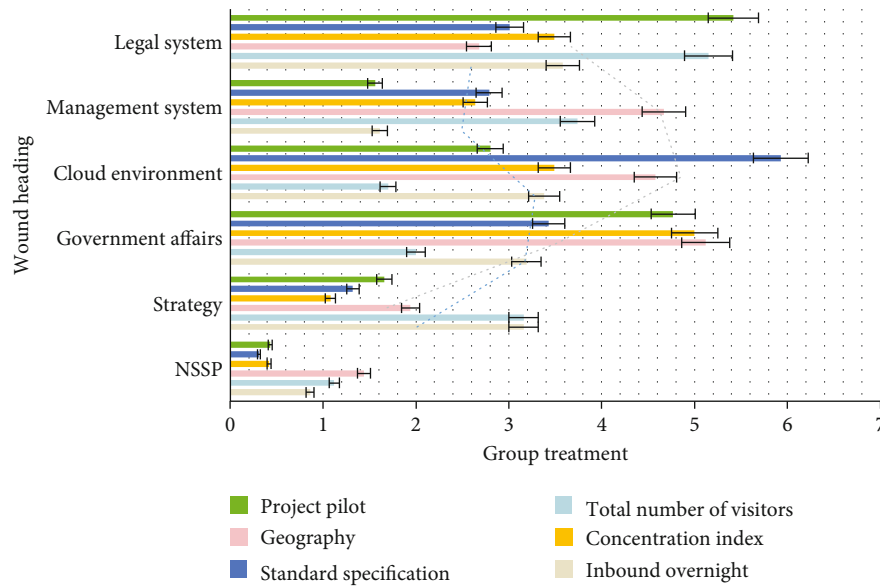


FIGURE 3: High-level rural tourism development potential.

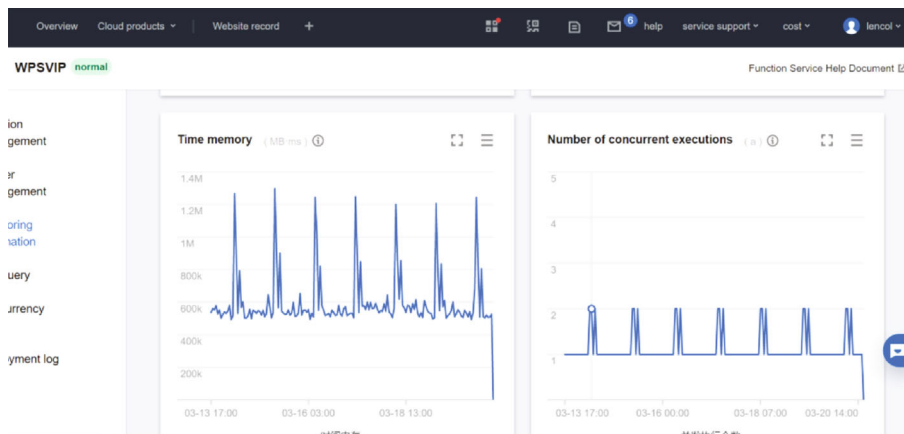


FIGURE 4: Tencent Cloud’s tourism data cloud computing data statistics chart (Statistics services provided by *Tencent Cloud* <https://console.cloud.tencent.com/>).

cloud security visually by producers, consumers, and decomposers. It also distinguishes the two kinds of tourism data in the cloud data Table 2.

As shown in Figure 5, the seasonal difference of regional passenger flow is obvious, and the regional passenger flow is mainly concentrated in summer and autumn. Generally speaking, the regional passenger flow presents the seasonal variation law. To explore the different characteristics, this study introduces the seasonal index and climate comprehensive comfort (CCI). The results show that the seasonality index of regional tourist flow tends to be 11.218 in recent years, and the seasonality index is too large and shows a stable trend in three years, which indicates that the seasonality difference of regional tourist flow presents a continuous solidification trend.

As shown in Table 3, the development potential of rural tourism in the region as a whole presents the characteristics

of “center periphery” spatial structure. The high level of rural tourism is concentrated in the provincial and municipal central city and its surrounding areas, while the low level of rural tourism development potential is scattered and far away from the provincial and municipal central city. However, the low-level radiation driving effect of high-level rural tourism development potential on the surrounding areas is not significant, and the overall regional linkage development and regional cooperation are not significant. It needs to be strengthened.

As shown in Figure 6, since the formal signing of the National Free Trade Zone framework, the number of national cloud computing guided tourism has been growing, reaching new heights. In terms of quantity, the total number of national tourists, inbound overnight tourists, and inbound foreign tourists in China is on the increase; in terms of proportion, the proportion of national tourists in the total number of

TABLE 2: Two scenarios of travel data to the cloud.

Item	Inbound overnight	Geography	Concentration index	Standard specification	Project pilot
NSSP	1.54	1.33	1.55	1.7	1.66
Strategy	2.22	3.22	2.04	2.84	2.31
Government	4.23	2.66	5.13	4.52	3.48
Management system	2.54	1.35	5.29	4.66	1.93
Legal system	2.25	2.01	2.89	1.31	4.64

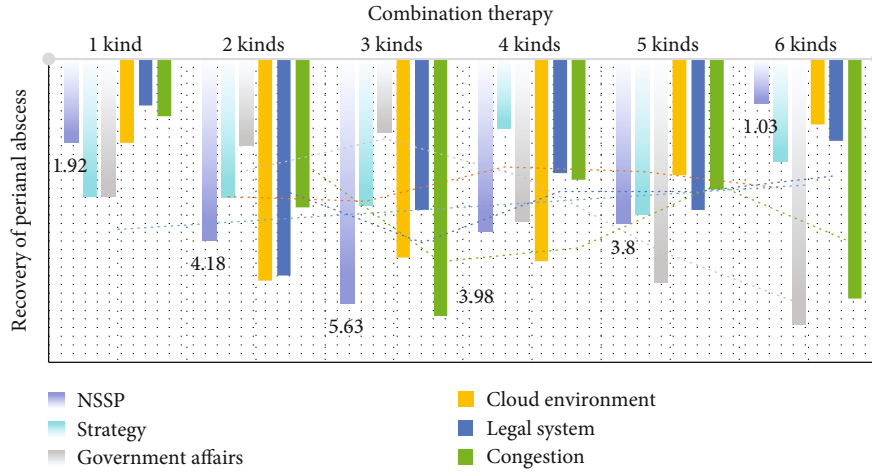


FIGURE 5: The seasonal difference of regional passenger flow is obvious.

TABLE 3: Regional rural tourism development potential.

Item	Management	Standard	Specification	Legal system	Market guidance
Center	1.44	0.5	0.61	1.35	0.92
Peripheral	1.37	2.91	1.93	1.67	2.71
NSSP	3.87	2.78	2.61	4.55	3.78
Strategy	3.94	1.69	3.34	1.75	2.48
Government	2.15	4.87	4.68	3.93	4.76

inbound overnight tourists and inbound foreign tourists is on a steady growth trend.

The geographical concentration index and share of the national cloud computing guided tourism market draw an interannual change chart, as shown in Figure 7. The geographical concentration index G of the main tourist source markets in China shows a continuous increasing trend, indicating that China’s foreign tourist source countries are more and more concentrated in geographical distribution. The value of G is between 0.9 and 2.33, and the degree of dependence on the national tourist market is relatively small. After 2019, the G value will rise to 9.60 in 2021, and China’s dependence on the national tourist market will also increase significantly. It can be seen that the status of the whole country in China’s rural leisure tourism market is increasingly important.

As shown in Figure 8, it is predicted that the cloud computing guided tourism demand of both the whole country

and the major source countries will basically maintain a sustained and stable growth trend, and the market development prospect is broad. In general, China is expected to receive 25.5799 million tourists from all over the country in 2026, far more than the total number of Asian inbound tourists (19.1207 million) in 2020, and it will remain at more than 25 million every year since then; among the seven major source countries, each source country has maintained a steady growth year by year. Therefore, with the increasingly close ties between China and the whole country, it can be predicted that the potential of the national tourist market is huge. In the future, it will not only continue to consolidate its market position as a major tourist source in Asia but also become a key area for the development of the rural leisure tourism market in China.

As shown in Figure 9, after ranking the correlation degree values, it is found that the influence degree of each factor on the national cloud computing guided tourism is

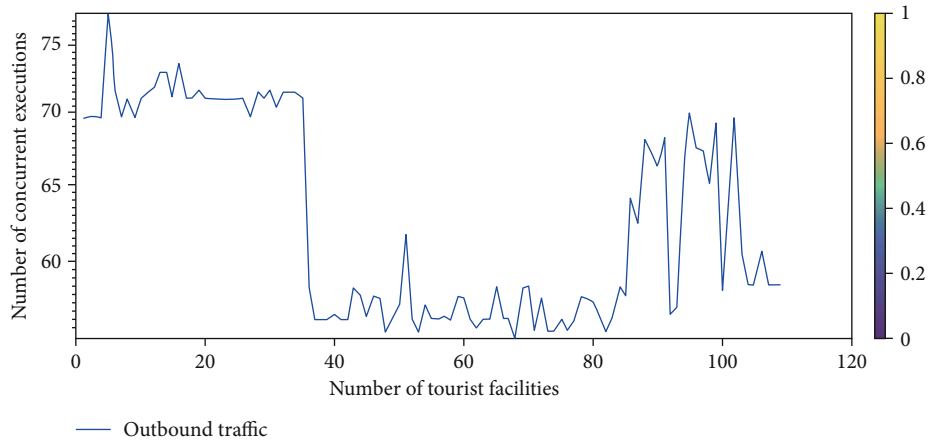


FIGURE 6: Cloud computing guides travel needs.

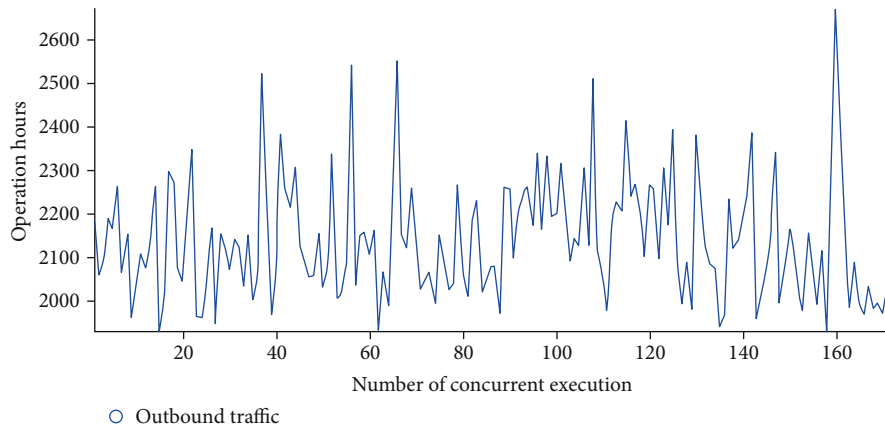


FIGURE 7: Cloud computing guides the number of people traveling.

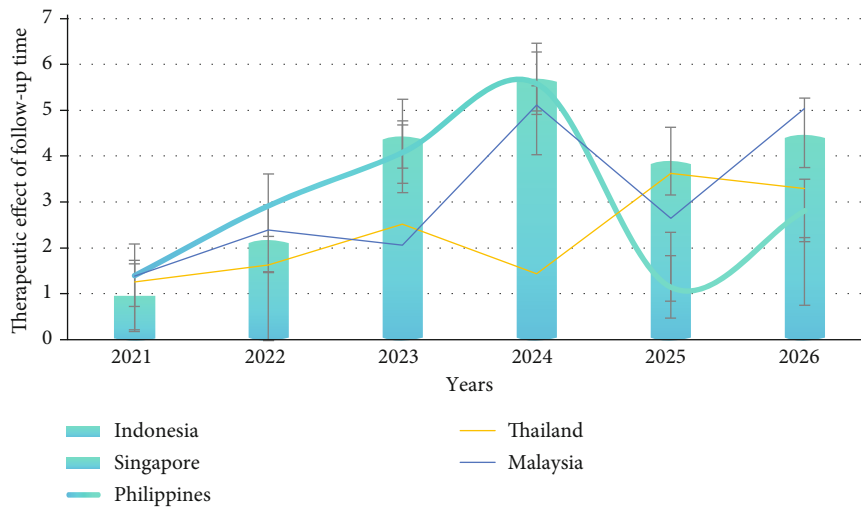


FIGURE 8: Cloud computing guides travel needs.

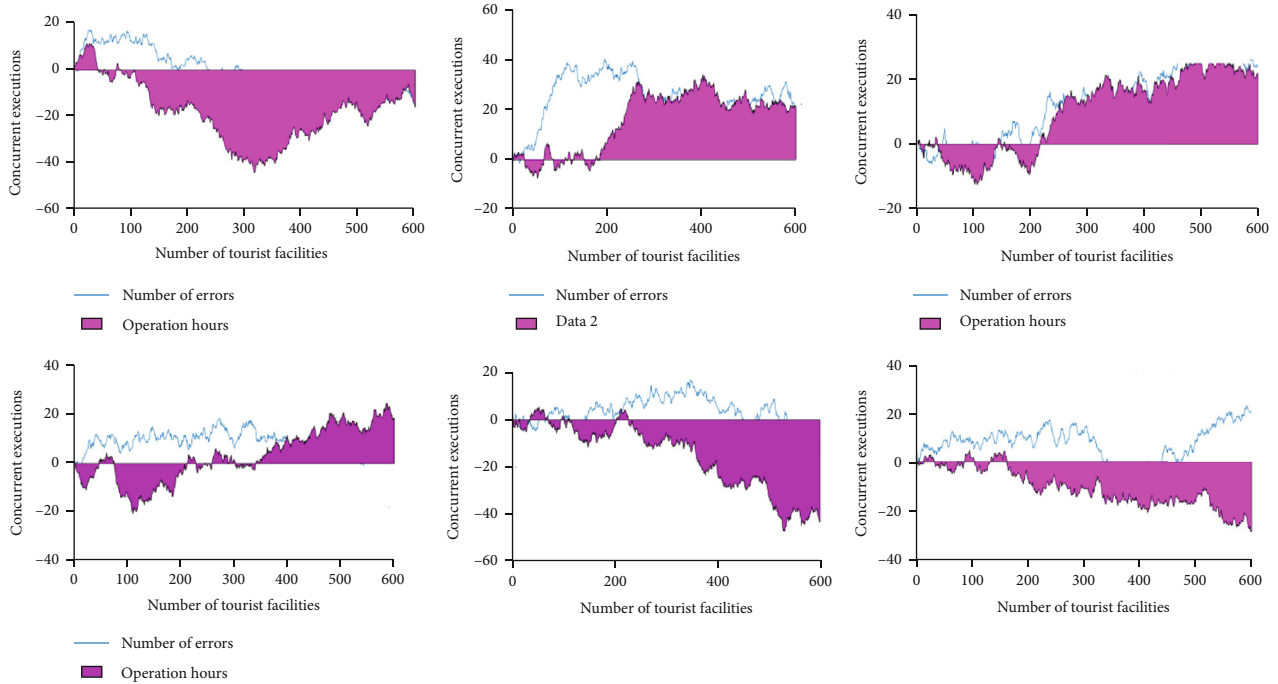


FIGURE 9: National cloud computing guide tourism market.

as follows: international route transportation mileage (0.832), bilateral trade volume (0.812), China's per capita GDP (0.805), tourism reception facilities (0.778), and the number of world heritage sites (0.765). Among them, the international route transportation mileage ($\times 4$) has the largest correlation, which indicates that for the national tourists, the tourism transportation convenience, especially the shipping conditions, has the most profound impact on their enthusiasm and demand in China. The correlation degree of bilateral trade volume ($\times 5$) takes the second place, which shows that under the background of globalization, the driving role of international trade in tourism cannot be underestimated. The higher the level of economic exchanges in China, the more effective it is to drive the demand of tourists to China.

4.2. Findings and Discussions. The statistical method of cloud computing is used to investigate the characteristics, development trend, and influencing variables of the national cloud computing tour guide industry. In terms of market development, the national cloud tour guide market exhibits an overall stable, continuous, and positive growth trend. The overall number of Chinese tourists is increasing, and rural leisure tourism in China has shifted to the national region. Rural leisure tourism is a significant component of the tourism industry. It has a general homogeneity and individual concentration in its temporal distribution, and the regional distribution of the interyear concentration index shows an increasing concentration trend. From 2019, the geographic concentration index begins to rise linearly, peaking at 9.60 in 2020.

At the same time, despite changes in country share rankings, market concentration is still high and dependence on a

few major source countries is too strong. We should take advantage of the "One Belt, One Road" initiative and China-FTA platform to strengthen mutual trust, cultural exchanges, and economic cooperation, accelerate the implementation of mutual visa-free tourism policy, promote tourism exchanges and cooperation, and improve the level of Mandarin reception to attract more tourists from all over the country. On the other hand, differentiated development should be carried out for various types of visitor source markets. The correlation between China's GDP per capita and resident rural leisure tourism is in the middle of the range, with no significant difference from the first two indicators, indicating that national economic development is closely related to resident rural leisure tourism, and national travelers are more willing to go to countries with better economic development. Domestic travelers are concerned about China's tourism resource endowment and tourism reception capacity, as the correlation between the number of tourism reception facilities and the number of world cultural heritage sites is also greater than 0.75.

5. Conclusions

The growth potential of rural tourism has shown apparent geographical autocorrelation and spatial agglomeration in the mobile cloud IoT computing environment. The counties with the most potential for rural tourist growth, as well as those with the least, exhibit geographical clustering patterns. The province's high-level rural tourist development potential is concentrated in the center, whereas the province's low-level rural tourism development potential is dispersed across the province, city, and district. The radiation impact of high-level regions on nearby low-level growth potential

areas, on the other hand, is minimal, and overall regional linkage development and regional collaboration must be increased. The growth potential of rural tourism in the mobile cloud IoT computing environment is unequal, and the difference is mostly due to the abundance of rural tourism resources and the degree of economic development.

Data Availability

All the data used is given in the paper.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] X. Chen, L. Jiao, W. Li, and X. Fu, "Efficient multi-user computation offloading for mobile-edge cloud computing," *IEEE/ACM Transactions on Networking*, vol. 24, no. 5, pp. 2795–2808, 2016.
- [2] Y. Dai, D. Xu, S. Maharjan, and Y. Zhang, "Joint computation offloading and user association in multi-task mobile edge computing," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 12, pp. 12313–12325, 2018.
- [3] Y. Zhou, L. Tian, L. Liu, and Y. Qi, "Fog computing enabled future mobile communication networks: a convergence of communication and computing," *IEEE Communications Magazine*, vol. 57, no. 5, pp. 20–27, 2019.
- [4] Y. Qi, L. Tian, Y. Zhou, and J. Yuan, "Mobile edge computing-assisted admission control in vehicular networks: the convergence of communication and computation," *IEEE Vehicular Technology Magazine*, vol. 14, no. 1, pp. 37–44, 2019.
- [5] Y. Zhou, L. Liu, L. Wang et al., "Service-aware 6G: an intelligent and open network based on the convergence of communication, computing and caching," *Digital Communications and Networks*, vol. 6, no. 3, pp. 253–260, 2020.
- [6] P. Porambage, J. Okwuibe, M. Liyanage, M. Ylianttila, and T. Taleb, "Survey on multi-access edge computing for internet of things realization," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 2961–2991, 2018.
- [7] S. K. Smith and M. House, "Snowbirds, sunbirds, and stayers: seasonal migration of elderly adults in Florida," *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, vol. 61, no. 5, pp. S232–S239, 2006.
- [8] E. Charles-Edwards and M. Bell, "Seasonal flux in Australia's population geography: linking space and time," *Population Space and Place*, vol. 21, no. 2, pp. 103–123, 2015.
- [9] D. A. Sullivan and S. A. Stevens, "Snowbirds," *Research on Ageing*, vol. 4, no. 2, pp. 159–177, 1982.
- [10] T. D. Hogan, "Determinants of the seasonal migration of the elderly to sunbelt states," *Research on Ageing*, vol. 9, no. 1, pp. 115–133, 1987.
- [11] J. Silaa, H. Jazri, and H. Muyingi, "A study on the use of mobile computing technologies for improving the mobility of Windhoek residents," *African Journal of Science Technology Innovation and Development*, vol. 9, no. 1, pp. 1–15, 2021.
- [12] Y. Li, K. Kim, and Y. Ding, "Early warning system of tennis sports injury risk based on mobile computing," *Mobile Information Systems*, vol. 2021, Article ID 3278276, 10 pages, 2021.
- [13] J. Zhang, X. Hu, Z. Ning et al., "Energy-latency tradeoff for energy-aware offloading in mobile edge computing networks," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2633–2645, 2018.
- [14] F. Wang, J. Xu, X. Wang, and S. Cui, "Joint offloading and computing optimization in wireless powered mobile-edge computing systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 1784–1797, 2018.
- [15] Y. Ding, C. Liu, X. Zhou, Z. Liu, and Z. Tang, "A code-oriented partitioning computation offloading strategy for multiple users and multiple mobile edge computing servers," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4800–4810, 2020.
- [16] G. Zhefeng, J. Wang, J. Congfeng et al., "Analysis of resource utilization of co-located clusters," *Chinese Journal of Computers*, vol. 43, no. 6, pp. 1103–1122, 2020.
- [17] C. Xu, X. Xiong, Q. Du et al., "Dynamic scheduling model of rail-guided vehicle (RGV) based on genetic algorithms in the context of mobile computing," *International Journal of Mobile Computing and Multimedia Communications*, vol. 12, no. 1, pp. 43–62, 2021.
- [18] A. Alkhalil, "Evolution of existing software to mobile computing platforms: framework support and case study," *International Journal of Advanced and Applied Sciences*, vol. 8, no. 3, pp. 100–111, 2021.
- [19] Y. Hmimz, T. Chanyour, M. el Ghmary, and M. O. Cherkaoui Malki, "Joint radio and local resources optimization for tasks offloading with priority in a mobile edge computing network," *Pervasive and Mobile Computing*, vol. 73, no. 1, article 101368, 2021.
- [20] J. He, D. Zhang, Y. Zhou, and Y. Zhang, "A truthful online mechanism for collaborative computation offloading in mobile edge computing," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4832–4841, 2020.
- [21] C. Wang, C. Liang, F. R. Yu, Q. Chen, and L. Tang, "Computation offloading and resource allocation in wireless cellular networks with mobile edge computing," *IEEE Transactions on Wireless Communications*, vol. 16, no. 8, pp. 4924–4938, 2017.
- [22] T. Wang, M. Z. A. Bhuiyan, G. Wang, L. Qi, J. Wu, and T. Hayajneh, "Preserving balance between privacy and data integrity in edge-assisted internet of things," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 2679–2689, 2020.
- [23] Y. Bowen, P. Lingjun, X. Yuting, X. Jingdong, and Z. Jianzhong, "Joint task offloading and base station association in mobile edge computing," *Journal of Computer Research and Development*, vol. 55, no. 3, pp. 537–550, 2018.
- [24] Y. Wang, M. Sheng, X. Wang, L. Wang, and J. Li, "Mobile-edge computing: partial computation offloading using dynamic voltage scaling," *IEEE Transactions on Communications*, vol. 64, no. 10, pp. 4268–4282, 2016.
- [25] T. Q. Dinh, "Offloading in mobile edge computing: task allocation and computational frequency scaling," *IEEE Transactions on Communications*, vol. 65, no. 8, pp. 3571–3584, 2017.

Research Article

Multitarget Real-Time Tracking Algorithm for UAV IoT

Tao Hong ^{1,2}, Qiye Yang ², Peng Wang,³ Jinneng Zhang,⁴ Wenbo Sun,⁵ Lei Tao,² Chaoqun Fang,² and Jihan Cao²

¹Yunnan Innovation Institute-BUAA, Kunming, China

²Beijing Key Laboratory for Microwave Sensing and Security Applications, Beihang University, Beijing, China

³Joint War College, National Defense University of PLA, Beijing, China

⁴Beijing University of Agriculture, Beijing 102206, China

⁵Aerospace Hi-Tech Holding Group Co. Ltd., Beijing, China

Correspondence should be addressed to Tao Hong; hongtao@buaa.edu.cn

Received 8 March 2021; Revised 23 June 2021; Accepted 17 July 2021; Published 26 August 2021

Academic Editor: Bo Rong

Copyright © 2021 Tao Hong 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.

Unmanned aerial vehicles (UAVs) have increased the convenience of urban life. Representing the recent rapid development of drone technology, UAVs have been widely used in fifth-generation (5G) cellular networks and the Internet of Things (IoT), such as drone aerial photography, express drone delivery, and drone traffic supervision. However, owing to low altitude and low speed, drones can only limitedly monitor and detect small target objects, resulting in frequent intrusion and collision. Traditional methods of monitoring the safety of drones are mostly expensive and difficult to implement. In smart city construction, a large number of smart IoT cameras connected to 5G networks are installed in the city. Captured drone images are transmitted to the cloud via a high-speed and low-latency 5G network, and machine learning algorithms are used for target detection and tracking. In this study, we propose a method for real-time tracking of drone targets by using the existing monitoring network to obtain drone images in real time and employing deep learning methods by which drones in urban environments can be guided. To achieve real-time tracking of UAV targets, we employed the tracking-by-detection mode in machine learning, with the network-modified YOLOv3 (you only look once v3) as the target detector and Deep SORT as the target tracking correlation algorithm. We established a drone tracking dataset that contains four types of drones and 2800 pictures in different environments. The tracking model we trained achieved 94.4% tracking accuracy in real-time UAV target tracking and a tracking speed of 54 FPS. These results comprehensively demonstrate that our tracking model achieves high-precision real-time UAV target tracking at a reduced cost.

1. Introduction

The application of 5G and the Internet of Things (IoT) represents the development of future drone technology. The development of 5G technology has facilitated the emergence of smart cities [1]. The recent construction of smart cities in China has been elevated to a national strategy with the strong support of the state, and considerable progress has been made. As of February 2019, 100% of subprovincial cities and 93% of prefecture-level cities of China—consisting of more than 700 cities in total (including county-level cities)—have proposed or constructed smart cities. Regardless of the scale of the smart city market or smart city information technology investment, both exhibit a rapid growth trend and a large

future market space. IDC (Internet Data Center), an international data company, predicts that investments related to global smart city technology will reach 189.46 billion US dollars in 2023. Meanwhile, the same investment in China will reach 38.92 billion US dollars.

Smart cities comprehensively promote the development of modern living through information technologies, such as the IoT, cloud computing, and geospatial infrastructure, in Figure 1. Cities that want high-quality development need smarter infrastructure, in addition to roads, viaducts, hydro-power, and so on. Smart city infrastructure, such as the IoT, tends to become increasingly popular with the development of 5G networks. Smart cities present new changes on multiple levels, and the construction of smart cities provides



FIGURE 1: Drone in IoT smart cities.

unprecedented opportunities for development [2]. To support the information infrastructure of smart cities, the focus is diverted to full coverage of the 5G network and object recognition cameras, among others, which can be gradually integrated with the IoT infrastructure to achieve a comprehensive IoT. From the perspective of technological development, smart city construction requires the realization of full perception, interconnection, universal computing, and integrated applications through the IoT, cloud computing, and other modern information technology applications represented by mobile technology.

The advantage of a smart city lies in its safety and security. The steady progress of 5G technology has also facilitated the continued optimization of city monitoring systems. Traditional security surveillance system cameras can be replaced by smarter 5G cameras. 5G technology has three characteristics: (i) high throughput, which solves the bandwidth transmission problem of video upstream and downstream; (ii) low-latency, which can achieve a theoretical value of 1 ms or 10 ms in the 5G era; and (iii) ultra-large-scale. 5G inherently supports ultra-large-scale device access and can support more cameras and other IoT devices for the security industry [3]. Only a city surveillance system with a wider bandwidth and higher video stability can be combined with an intelligent video surveillance cloud platform to achieve true intelligent security. Moreover, compared with wired video transmission, 5G wireless transmission is easier to deploy, more convenient, and lower in cost. The high transmission, high broadband, and high reliability of 5G can provide a UAV identification system with more high-definition monitoring data at a faster speed. Using a surveillance system composed of a network of urban surveillance cameras to monitor, identify, and warn drones in designated areas in real time can effectively solve the problems in drone surveillance. The key to this method is to determine how to effectively detect whether a drone to track and locate from the video exists. In this study, we attempt to solve this problem by machine learning.

With the continuous industrial and technological progress, low-altitude, slow-speed, and small-target UAVs have

been rapidly developed and are widely used. UAVs fly at low altitudes; thus, radar waves may not reach these targets as they are affected by the curvature of the earth and the shelter of buildings. In addition, a large amount of ground clutter will enter the radar receiver while it is working, which impedes radar from distinguishing the echo signal of a UAV target. The flying speed of UAVs is slow, and some are even lower than the radar speed detection threshold, preventing the pulse-Doppler radar from detecting the target. Slow-flying UAVs are easily confused with slow-moving clutter, such as weather clutter and bird swarm, rendering target t recognition difficult. Their small size and radar reflection area, in addition to the weak echo signal, weaken w the UAV detection ability of radar. However, owing to difficulties in target detection and effective supervision, such UAVs have incurred major security threats to countries in recent years. In 2017, UAV interference occurred at Kunming Changshui International Airport in China, resulting in 35 flights being forced to divert to alternate routes and 28 flights being delayed at the airport. Some flights were delayed for 4 h, and the airport runway was forced to close for 45 min. In 2018, “black flight” and “disturbing flight” UAV events in Germany showed rapid growth trends. As of August 12, 2018, more than 100 UAV interference incidents occurred in major airports in Germany, which exceeded the total number reported in 2017. Thus, methods for detecting such targets have significant and immediate application requirements. Various recognition methods for UAV detection currently exist, including radar technologies, audio signal analysis, trajectory analysis, and image recognition.

Radar technology has been widely used in UAV detection and classification because of its fast-remote sensing ability. In radar technology, frequency-modulated continuous-wave (FMCW) radar is the most common choice because it can obtain sufficient information about targets with a short dwell time. When the position of the UAV target changes relative to the radar, the rotation of the wing can cause the radar echo to modulate and produce a micro-Doppler effect. Analysis of the micro-Doppler characteristics of radar echo can extract detailed information such as the number of rotors of the UAV target. The use of FMCW in surveillance systems is limited because it cannot detect the distance from the target to the radar, requiring manual intervention [4]. For UAV recognition based on audio signals, the multirotor UAV can produce FM (Frequency Modulation) noise when flying. The current study proposes the RPM (rotation per minute) speed wave correction method, which corrects the classical noise prediction method of rotorcrafts, considers the frequency modulation effect, and identifies UAVs. Large-scale public activity venues, airports, government agencies, and other places requiring UAV detection have large background noise; thus, the noise emitted by UAVs for identification is difficult to capture [5]. To address this concern, a UAV flight path recognition network based on radar monitoring data is constructed in accordance with the conditions identified by research on UAV flight path identification and the characteristics of radar data information of low-altitude surveillance systems. A UAV identification method based on radar monitoring data and recurrent neural network (RNN) flight path

identification network is then proposed. Modeling with RNN-Long Short-Term Memory networks after training can identify the target trajectory in airspace, efficiently identify the trajectory data of the UAV, and achieve the expected classification effect. However, collecting trajectory information entails time, suggesting that a UAV is impossible to identify and track in real time.

Traditional UAV detection techniques are often limited in public places, government agencies, airports, and other areas. Moreover, radar systems experience difficulty operating effectively in environments with many shelters, where the RPM cannot operate because excessive background noise and track recognition cannot meet the requirements of real-time monitoring. For the past few years, the rapid improvement of deep learning and computing abilities have greatly enhanced the accuracy and speed of image target recognition and classification, allowing the tracking of algorithms based on target detectors. Relevant research has determined that existing methods employing a target detection algorithm to identify UAVs can only be used to identify targets but not to target the identified drones. Meanwhile, the existing UAV tracking algorithm has deficiencies in robustness and real-time performance. To address these issues, this study introduces the target tracking algorithm into the UAV recognition algorithm.

2. Related Work

The basic idea of multitarget real-time tracking for UAV positioning in IoTs is illustrated in Figure 2. The IoT camera is arranged in the city, and the captured images are transmitted to the cloud via 5G at a high speed. The real-time machine learning algorithm is used to track multiple targets of the drone. To achieve real-time target tracking, research on related machine learning algorithms is also necessary.

Currently, the most widely used deep learning target detection method is the target detection algorithm based on CNNs (Convolutional Neural Networks). The development of CNNs has a long history. In 1962, Hubel and Wiesel used the brain of a cat to explore the visual system. In 1980, Japanese scientist Kunihiko Fukushima proposed a neural network structure with a convolutional layer and a pooling layer. In 1998, Yann LeCun proposed LeNet-5 and applied the backpropagation algorithm for training this type of neural network structure, forming a prototype for the contemporary CNN. In the 2012 ImageNet Image Recognition Competition Challenge, AlexNet proposed a deep structure and dropout method discussed in the study by Hinton et al. in which the error rate decreased from more than 25% to 15%, revolutionizing the field of image recognition. Following the idea of AlexNet, LeCun et al. proposed DropConnect in 2013, further reducing the error rate to 11%. Network in Network was proposed by Yan Shuicheng et al. of NUS (National University of Singapore), significantly altering the structure of CNN. Based on these methods, Inception and VGG architectures deepened the network in 2014 to about 20 layers and significantly improved the image recognition error rate to 6.7%, indicating its similarity to the human error rate of 5.1%. Ren Shaoqing and He Kaiming et al. of MSRA

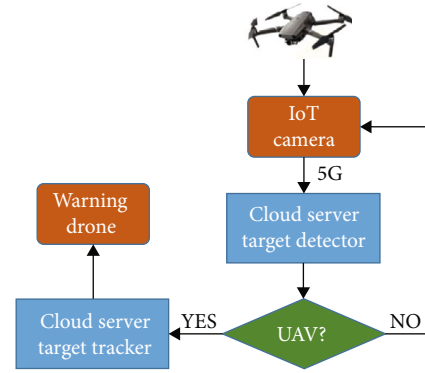


FIGURE 2: Drone tracking flowchart.

(Microsoft Research Asia) optimized the original R-CNN (region-based CNN) and Fast R-CNN with the development of Faster R-CNN. The main contributions of Faster R-CNN are the use and image recognition of the same CNN features in which features can not only recognize the category of objects in the image but also record their positions. He Kaiming subsequently introduced Mask R-CNN and added a mask head to Faster R-CNN. By using the mask head only in training, the mask head message was passed back to the original CNN feature, allowing the original feature to contain more detailed information. Currently, the structure of CNN is becoming increasingly complex, and the research direction is to find algorithms that can automatically optimize its structure. Supported by a deep coproduct neural network, the target detection method can be divided into two categories: the two-stage method and the one-stage method [6–8].

2.1. Target Detection Algorithms

2.1.1. Two-Stage Approaches. The two-stage approach divides the target detection task into two phases: RoI (region of interest) extraction, followed by RoI classification and regression. R-CNN, SPPNet, Fast R-CNN, Faster R-CNN, Mask R-CNN, and Cascade R-CNN, among others, are all two-stage approaches. Fast R-CNN before object detection is driven by the region proposal approach and the region-based CNN. Fast R-CNN uses a very deep network to achieve near real-time rates when the time spent on regional proposals is ignored; meanwhile, the proposal is also the bottleneck of inference in the detection system. Regional proposal methods often rely on inexpensive functions and economic reasoning schemes. Selective search is a widely used technique but is slower than effective detection networks by about one order of magnitude [9].

2.1.2. One-Stage Approaches. The one-stage method eliminates RoI extraction and then directly classifies and regresses the candidate anchor boxes. As its name suggests, this approach follows a completely different objective to apply a single neural network to the entire image. The representative algorithms are YOLO, R-SSD, RefineNet, and so on. Contrary to the classifier-based method, YOLO is trained on the loss function directly corresponding to the detection performance, while the whole model is trained under joint training [10]. As a representative one-stage method, YOLO has

significant advantages in computational speed over the two-stage method. Subsequent versions of YOLO (YOLOv2 and YOLOv3) exhibit improved detection accuracy while maintaining a high detection speed.

The development of anchor-based approaches has involved several issues. In the anchor mechanism, solid prior knowledge is needed to set the appropriate super parameters of scale and aspect ratio, and the number of targets in an image is often limited. Setting a large number of anchor boxes based on each anchor can produce a large number of easy samples and lead to an out-of-balance state of positive and negative samples. When target classification is based on the anchor box, the threshold of intersection over union (IOU) is also difficult to set. By contrast, the anchor-free method can avoid the anchor. While it cannot provide high detection stability, its calculation time is significantly reduced, allowing real-time high-precision detection and segmentation.

2.1.3. Anchor-Based. These methods typically require numerous anchors to ensure a sufficiently high IOU rate on the ground. Various hyperparameters and design choices are also possible with anchor boxes, and these methods can become more complex when these choices are used in conjunction [11].

2.1.4. Anchor-Free. Generally, the nonanchor detector belongs to a first-class detector. Although the performance of the one-stage approach CornerNet remains limited by its relatively weak ability to reference global information, its productivity can be increased by its ability to perceive visual patterns within each proposed area. Thus, it can independently identify the correctness of each bounding box.

2.2. Target Tracking Algorithms. Early tracking patterns used temporal and spatial points of interest, which could not be separated from some low-level features, such as corners and intensity peaks [12, 13]. Although early classical algorithms were able to achieve high-precision single-target tracking, they could not meet the high-speed multitarget tracking problem. Subsequently, because of the rising interest and financial support of target detection in recent years [14], detection by tracking has gradually led the research and application of multitarget detection. The current target tracking algorithm completes detection and tracking simultaneously, transforming the existing detector into a tracker.

2.2.1. Classic Algorithms. The improved classic MHT (Multiple Hypothesis Tracking) algorithm of the 1990s has achieved performance close to that of the most advanced methods in the standard benchmark dataset [15]. Using accurate object detectors simplifies a small number of possible assumptions. The appearance model can be learned efficiently using a regularized least-squares framework, and each assumed branch requires only several additional actions to take advantage of MHT when using higher-order information. JPDA (Joint Probabilistic Data Association) has also been improved to compute for high target and clutter density applications. Many experiments have also shown that when embedded in the simple tracking framework, the JPDA algorithm performs competitively, with the most advanced global tracking methods in both applications, while significantly

reducing the processing time [16]. While MHT and JPDA retain their advantages, uncertainty and latency are high when they are faced with difficult targets. The composite complexity of these methods increases exponentially with the number of trace lines, making it impractical to apply such traditional methods in highly dynamic environments.

2.2.2. Multitarget Tracking. The latest research trend in multitarget tracking is the use of the same framework to include detection and tracking and the combination of the previous two-stage methods into a one-stage multitasking process. This technique uses the current and previous frames as input and then predicts the target frame offset position of the next frame. CenterTrack applies the detection model to a pair of images and detects from previous frames. With minimal input and the previous frame, CenterTrack can locate objects and predict their association. Simple, online, and live, CenterTrack can also be easily extended to single-eye 3D tracking by tracking other 3D attributes [17]. The detector based on the keypoint has shown satisfactory performance. However, incorrect key matching still commonly occurs and can seriously inhibit the performance of the detector. Compared with traditional embedding methods, CentripetalNet combines location information with match corner points more accurately. The corner pool extracts the information from the bounding box to the boundary. Feature matching of the star-shaped deformable convolution network endows the corner information with clarity. In addition, by equipping CentripetalNet with a mask prediction module, it can explore instance segmentation on an anchor-free detector [18].

2.2.3. Tracking by Detection. Mainstream target tracking algorithms are currently designed according to the detection and tracking mechanism. First, the target detection algorithm is used to detect the target in each frame and obtain the corresponding position coordinates, classification, confidence, and other data. Data association is then employed to correlate previous detection results with the detection results of the previous frame. However, when the target moves rapidly, traditional tracking by detecting matching fails. The tracking method based on trajectory prediction can successfully solve this problem. By adding a one-step Kalman filter to predict the tracking state of the next frame, the predicted target state is compared with the detected target to link the fast-moving objects.

Compared with detection and tracking combined into multitasking, the traditional tracking-by-detection tracking algorithm requires no high-cost video streaming data training and only needs to make several improvements on the basis of the existing target detector and thereby obtain robust and real-time compliance tracking.

The advantage of the algorithm used in this study is that it does not rely on the video streaming dataset to train the model. Only image datasets can be used to train effective target tracking models to realize real-time tracking of UAV models.

3. Detection and Tracking Method of UAV

This study primarily is aimed at achieving a simple and effective drone tracking method. In the target detector

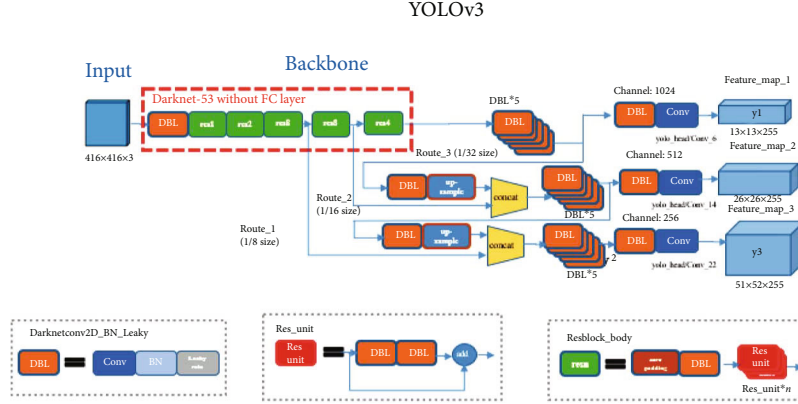


FIGURE 3: YOLOv3 network structure.

component, YOLOv3 [19] and CenterNet are selected to satisfy the speed and performance requirements of the target tracking model. As a representative of a single-stage network, YOLOv3 has no complicated network structure and can be well applied in industrial landing. As a representative of the new-generation anchor-free detection model, CenterNet can provide a good comparison with the algorithm under the anchor mechanism. The strategy in this study is to add data association between image frames on the existing target detector to achieve target tracking. Moreover, the most current detection and tracking methods require video streaming data as training data. However, the dataset entails considerably high costs. Consequently, we abandoned this tracking algorithm and turned to the mainstream tracking-by-detection tracking algorithm. This approach provides a clever strategy to complete the training of high-precision detection and tracking models with only image datasets.

3.1. Detector

3.1.1. Target Detector YOLOv3 and Its Improved Design. YOLOv3 is the third-generation version of the YOLO series. Compared with the two previous versions, the third-generation version integrates newly proposed techniques in target detection. The backbone partly draws on the structural design of the residual network. In the Darknet-53 [20] network, the neck part introduces the feature pyramid network (FPN) [21] structure. The superior techniques of the previous version are simultaneously retained, such as batch normalization (BN) [22] and k -means clustering algorithm [23]. The network structure of YOLOv3 is presented in Figure 3.

In YOLOv3, the network uses an improved backbone, upgrading Darknet-19, which was used by YOLOv2, to Darknet-53. Darknet-53 uses a 3×3 convolution kernel to replace the pooling layer, cuts the feature map, reduces the dimensionality, and performs a 1×1 convolution operation on the cut result to control the final output channel number. This approach not only reduces the amount of data brought by pooling and accelerates the calculation but also increases the nonlinearity and robustness of the network.

Recognizing objects at different scales is a basic challenge of Computer Vision, hence the proposal of the FPN struc-

ture. YOLOv3 introduces a structure similar to that of FPN to perform information fusion on detection frames with three different sizes. The FPN structure is aimed at using the inherent multiscale pyramid layer of the deep CNN to construct the feature pyramid, and its implementation combines low-level features with enhanced resolution and positioning information, together with high-level features with strong semantic information. This approach helps the network in detecting multiscale targets, particularly small ones. The information fusion of FPN includes three routes. The first route is a bottom-up pathway, which generates multiple feature maps of different layers through the forward propagation of the backbone CNN. For the feature pyramid, each stage is defined as a pyramid level. It is a top-down pathway, which is the core of the FPN structure, as shown in Figure 4. Its main function is the upsampling of the high-order pyramid layer to the same size as that of the secondary feature map. The lateral connection, which is mainly the high semantic feature map and high positioning information feature map obtained from the previous two roads, is combined by interpolation. Three sizes of prediction results— 13×13 , 26×26 , and 52×52 —are output to the network.

To accelerate the convergence speed during network training and prevent gradient explosion or disappearance in the backpropagation of the deep network, BN technology was integrated with YOLOv2 and retained in YOLOv3. In training the neural network, the input distribution of each layer of the network is always changing at each stochastic gradient descent for each minibatch. Owing to the sensitivity interval of the activation function being fixed, when continuous multilayer input distributions are present, the input distribution is in the activation function. In nonsensitive areas, network convergence tends to stall. To solve this problem, BN is added. The BN operation restricts the input distribution to a standard normal distribution with a value of 0 and a variance of 1 via a normalized operation; thus, the gradient change moves toward the optimal value of the loss function. This approach accelerates network convergence.

YOLOv3 is a representative of the anchor mechanism in the single-stage detection model. The method of setting set a suitable anchor is a finishing touch to enhance the detection performance of the model. In the anchor setting of YOLO, k

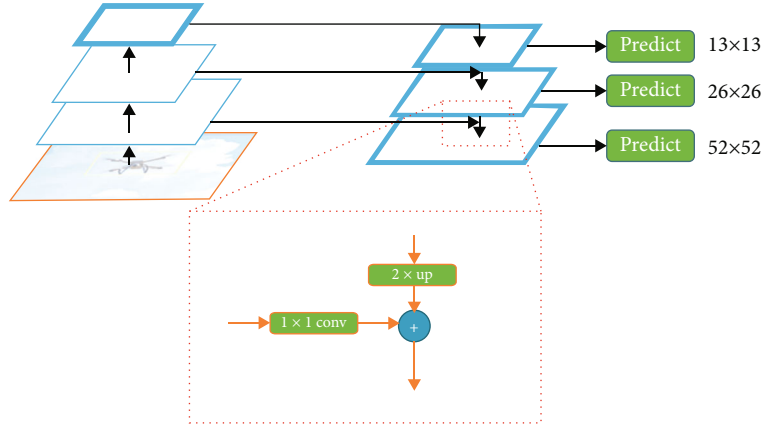


FIGURE 4: YOLOv3 feature pyramid network structure.

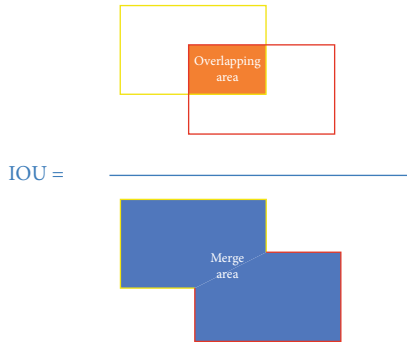


FIGURE 5: IOU calculation diagram.

-means clustering is still used for data processing to obtain the distribution setting of the anchor. Unlike that in Faster R-CNN, the anchor is manually generated through a preset scale and coordinate relationship. In YOLOv3, cluster analysis needs to be performed on the bounding boxes in the training set to generate appropriately sized anchors. In this study, we use a modified k -means clustering algorithm.

Performing standard k -means clustering on the anchor can induce the large box to generate a larger error during clustering, and this error is expected to be almost unrelated to the size of the box. The key to solving this problem is the measurement of the distance between different anchors. We use the maximum IOU (intersection over union) to replace the Euclidean distance measure in the standard k -means clustering algorithm. The IOU calculation is presented in Figure 5. The formula is as follows:

$$d(\text{box, centroid}) = 1 - \text{IOU}(\text{box, centroid}). \quad (1)$$

The brief process of k -means clustering for anchor boxes is summarized:

- (i) The coordinates of all labeled ground truth frames in the training set are extracted
- (ii) The coordinates of all ground truth (GT) boxes are converted into the height information of the boxes

- (iii) K GT boxes are randomly selected as anchor boxes; with these k anchor boxes and the remaining GT boxes, IOU is determined and $d = 1 - \text{IOU}$ is calculated
- (iv) The GT boxes are classified. The distance $\{d(i, 1), d(i, 2) \cdots d(i, k)\}$ of each GT box is compared with each anchor box, and the smallest distance d is selected. The GT box owned by the k anchor boxes is ultimately recorded
- (v) The anchor box is updated. For each anchor box, the average value of the frame height of its GT box is calculated, and the value is used as the new size of the anchor box
- (vi) Steps (i) to (iv) are repeated until the size of the anchor box no longer changes

To perform the k -means clustering operation on the anchor section, we set $k = 9$ because YOLOv3 has three feature maps. Each feature map requires three anchors of different scales. To perform k clustering training on the anchor data information in the training set and then ultimately obtain nine anchor scales, (36, 17), (54, 28), and (81, 40) are first used on the smallest 13×13 feature map. For the 26×26 feature map of the intermediate size, the three-scale anchors (123, 71), (156, 128), and (259, 111) are used, and (206, 172), (275, 217), and (380, 283) are prepared for the largest 52×52 feature map.

The SPP-Net [24] network is used to improve the YOLOv3 network. After the first layer feature map of YOLOv3, a structure similar to the SPP network is added. This addition is intended to effectively avoid the problems of image area clipping and image distortion caused by zoom operations and to solve the problem of CNN for image repeated feature extraction, which greatly improves the speed of generating candidate frames and reduces computational costs. Adding the SPP module to YOLOv3 strengthens the feature pyramid and semantic information and realizes the extraction of local and global features on the smallest feature map, which is favorable for addressing the problem of large differences in target size. The structure of YOLOv3-SPP is presented in Figure 6.

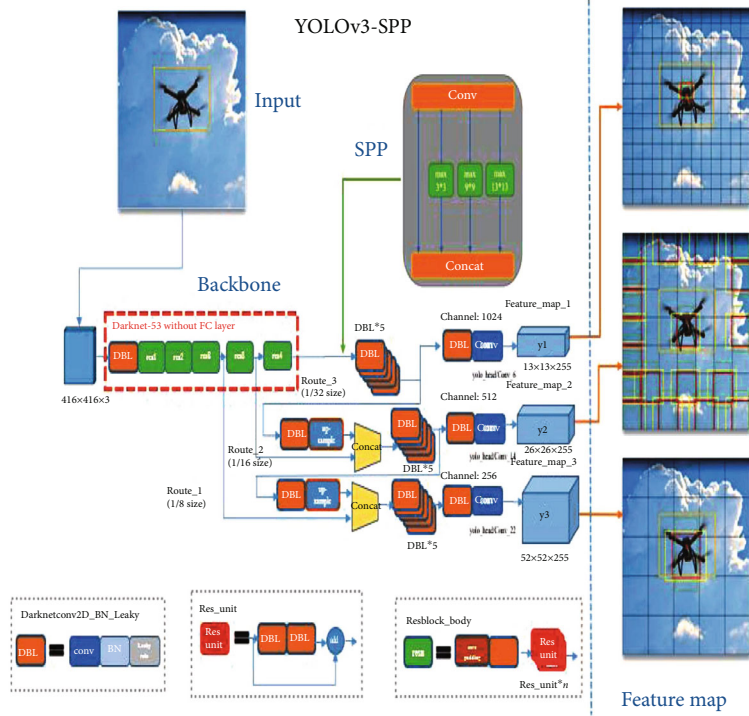


FIGURE 6: Modified YOLOv3-SPP network structure.

The loss function of the YOLO model must then be modified. The modified YOLO loss function mainly consists of box loss, confidence loss, and category loss:

$$\begin{aligned}
 \lambda = & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B T_{ij}^{\text{obj}} [(x_i - x_i)^2 + (y_i - y_i)^2 + (w_i - w_i)^2 \\
 & + (h_i - h_i)^2] + \sum_{i=0}^{S^2} \sum_{j=0}^B T_{ij}^{\text{obj}} (C_i - C_i)^2 \\
 & + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B T_{ij}^{\text{obj}} (C_i - C_i)^2 \\
 & + \sum_{i=0}^{S^2} T_i^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - p \wedge_i(c))^2.
 \end{aligned} \quad (2)$$

T_i^{obj} defines whether there is a target in the i -th grid; T_{ij}^{obj} defines whether the j -th *a priori* box in the i -th grid is the *a priori* box responsible for target prediction; and λ_{coord} is the frame loss. The coefficient λ_{noobj} is the coefficient of no target confidence loss; s is the side length of the feature map; and B is the number of anchors for each grid. The original YOLO loss function divides the width and height of the detection frame with a root sign and then calculates the loss. The purpose is to reduce the proportion of the width and height loss occupied by the target frame size and prevent the coordinate loss of the central point from being overwhelmed. However, in the k mean value after clustering, the loss of the target frame size is reduced. Consequently, the square root operation is no longer performed on the loss of the target frame

size. With this approach, the increased accuracy of anchor initialization is ensured, and the initial error exerts less effect and is easier to converge. The final loss category uses a cross-entropy loss function design.

3.1.2. Target Detector CenterNet and Its Improved Design. In recent years, the research on object detection has been redirected improving the detection accuracy of target detectors to focusing on the speed of the detector and then to the current weighting of the accuracy and real-time performance of the target detector. In this process, the two-stage detector method—for instance, the first stage of Faster R-CNN—proposes possible regions of interest (RoIs) via the region proposal network (RPN), combining the screening and calibration of RoIs. The first one is similar to the effect of coarse classification + fine classification. Owing to this design, Faster R-CNN achieves high detection accuracy, but simultaneously, the two-stage network runs at a low speed (5 FPS) and produces larger memory occupation, hindering its application in many real-world scenarios. The two-stage method is a slow process and thus has been reduced to a one-stage technique. The detection speed of one-stage detectors represented by SSD [25] and YOLO can generally be increased to more than 30 FPS, which basically meets the real-time requirements of detection and also achieves good detection accuracy. The most crucial problem at present is the poor detection efficiency of small target objects in multiscale detection tasks. Since the one-stage method does not have the same proposal stage as the RPN network layer, only a small part of the boxes can be selected as anchors when the anchor is preset, and few boxes match small objects. Many one-stage solutions have also been proposed in response to

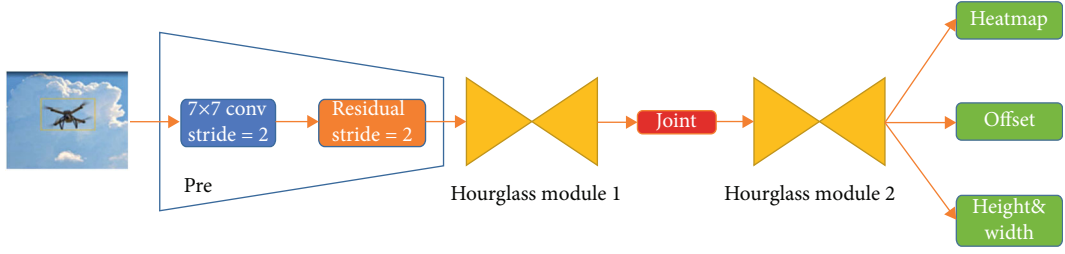


FIGURE 7: CenterNet network structure based on Hourglass backbone.

this situation. For instance, in the YOLO series, a k -means clustering algorithm is used for the dataset to generate a pre-set anchor. Moreover, an FPN-like structure is added to the train to enhance the positioning information of each layer, including semantic information. Since the development of target detectors, the anchor-based detection mechanism has gradually hindered the improvement of the detection performance. Therefore, a new method of solving object detection has emerged using the anchor-free technique.

The anchor-free method was originally proposed by CornerNet. Different from their previous target detector, the anchor is preset to predict the offset in order to complete the regression correction of the target frame. CornerNet directly converts the object detection problem into a keypoint detection problem. The network directly predicts the coordinates of the upper left point and the lower right point of the target frame and uses the embedding vector to match them and thereby complete the target detection. This approach abandons the anchor mechanism and no longer requires anchor setting; in addition, no subsequent screening and nonmaximum suppression (NMS) operations of a large number of anchors are needed. However, CornerNet also has deficiencies because some errors inevitably occur when the upper left corner and the lower right corner of the target box are matched, resulting in the reduced accuracy of the network. Moreover, the corner pooling it uses only relies on the edge information of the object and consequently uses no internal information, weakening its capability for global information acquisition. Therefore, we used CenterNet, a more advanced anchor-free detector (Figure 7).

CenterNet is further simplified in the idea of CornerNet. The network directly predicts whether each pixel is the target center; if so, it predicts the bounding box for the central point. This method is closely related to the anchor-based technique because each pixel in the feature map A pixel can be regarded as a shape-agnostic anchor. However, the “anchor” is only related to the position, and further predicting the offset and using NMS for postprocessing is unnecessary.

Among Figure 7, Pre is a 7×7 residual error unit with a step size of two. After this structure, the size of the picture is compressed to 1/4 of the original picture, and two hourglass modules perform keypoint detection. The final output has three branches: (i) Heatmap with the dimensions $(W/4, H/4, C)$, which outputs the center point positions of the objects in C categories; (ii) Offset, with the dimensions $(W/4, H/4, 2)$, which supplements and corrects the output result of Heatmap to improve the accuracy of positioning; and (iii) Height & Width, with the dimensions of $(W/4,$

$H/4, 2)$, which predicts the width and height of the detection frame centered on the keypoint.

The loss function of CenterNet also consists of three parts (target category loss, target center point offset loss, and target frame size loss):

$$L_{\text{det}} = L_k + \lambda_{\text{size}} L_{\text{size}} + \lambda_{\text{off}} L_{\text{off}}, \quad (3)$$

$$L_k = \frac{-1}{N} \sum_{xyc} \begin{cases} \left((1 - \hat{Y}_{xyc})^\alpha \log(\hat{Y}_{xyc}) \right), & \text{if } Y_{xyc} = 1, \\ \left((1 - Y_{xyc})^\beta (\hat{Y}_{xyc})^\alpha \log(1 - \hat{Y}_{xyc}) \right), & \text{otherwise,} \end{cases} \quad (4)$$

$$L_{\text{off}} = \frac{1}{N} \sum_p \left| \hat{O}_p - \left(\frac{p}{R} - \tilde{p} \right) \right|, \quad (5)$$

$$L_{\text{size}} = \frac{1}{N} \sum_p \left| \hat{S}_{p_k} - s_k \right|. \quad (6)$$

L_k is the target category loss function. For a $W \times H$ picture $I \in R^{W \times H \times 3}$, the keypoint heatmap generated after going through the network is $\hat{Y}_{xyc} \in [0, 1]^{(W/R) \times (H/R) \times C}$; R is the output stride; C is the number of keypoints; and N is the number of keypoints in the heatmap. In the heatmap, $\hat{Y}_{xyc} = 1$ indicates that the point is the center point of the target; $\hat{Y}_{xyc} = 0$ indicates that the point is the background; Y_{xyc} is the labeled GT information; α and β are the hyperparameters of focal loss [26]; and the default settings are 2 and 4. The category loss function in this part draws on the idea of focal loss. The central point of the training weight in the easy example is appropriately reduced, which is the loss value. When $Y_{xyc} = 1$, $(1 - Y_{xyc}^\alpha)$ acts as a correction function; if \hat{Y}_{xyc} is close to 1, point detection is relatively easy, and $(1 - Y_{xyc}^\alpha)$ is correspondingly lower. When \hat{Y}_{xyc} is close to 0, the proportion of training should be increased because the center point has not been learned. Therefore, $(1 - Y_{xyc}^\alpha)$ tends to be considerably large. The purpose is to balance the training process, that is, to solve the problem of imbalance between positive and negative samples. L_{off} is the offset loss of the center point, and $\hat{O}_p \in R^{(W/R) \times (H/R) \times C}$ is the local offset of each prediction center point, the target center point. The bias function must account for the accuracy loss of the input picture after the downsampling operation of the backbone network to render the prediction result closer to the target center point. In

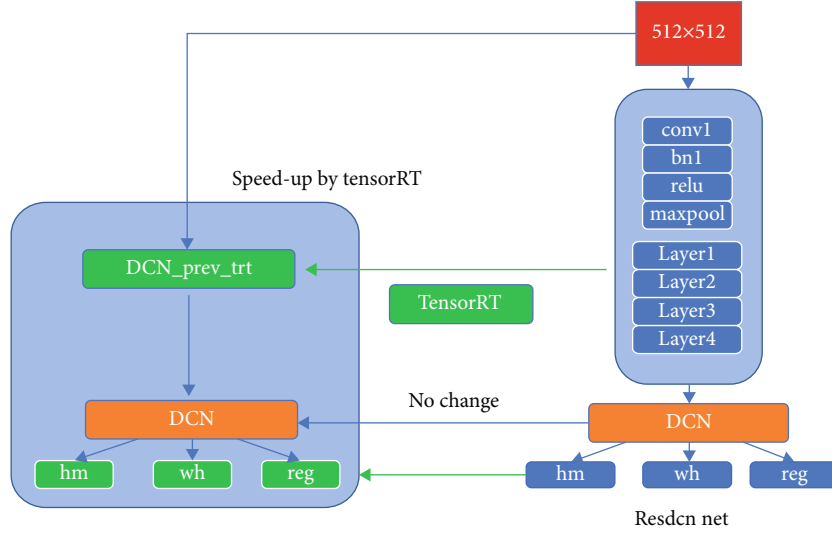


FIGURE 8: Split acceleration design.

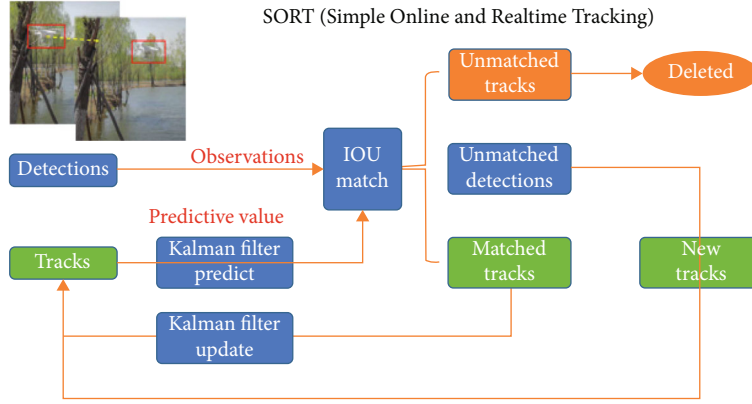


FIGURE 9: SORT algorithm flow.

this case, L_{size} is the loss function of the width and height of the target frame for the frame k with coordinates $(x_1^k, y_1^k, x_2^k, y_2^k)$. The size of the target frame $s_k = (x_2^k - x_1^k, y_2^k - y_1^k)$ and the loss function of the center point error are then designed by the loss function of L1 loss.

The backbone network of CenterNet consists of ResNet [27] and DCN [28]. However, upsampling convolutional layers, such as DCNv2, greatly limit the deployment of the model. To improve the performance of the CenterNet network and enable it to achieve real-time detection, we perform TensorRT acceleration processing on CenterNet. Limited by the design of the DCNv2 network, TensorRT does not currently support the acceleration operation of the DCN network. The network used appears in Figure 8.

The CenterNet network is separated to achieve TensorRT acceleration. The network is divided into two parts of the backbone network, ResNet and DCN, and ResNet is separately accelerated using TensorRT. This design can theoretically increase the running time of the backbone network by 10–20 times.

3.2. Tracker. After determining the detectors YOLOv3-SPP and CenterNet, we use Deep SORT [29] as the follow-up tracking algorithm. The Deep SORT algorithm is improved on the basis of the Simple Online and Realtime Tracking (SORT) algorithm [30].

The core of the SORT algorithm presented in Figure 9 consists of the Kalman filter and the Hungarian algorithm. The Kalman filter algorithm is divided into two processes: prediction and update. The algorithm defines the motion state of the target as eight normally distributed vectors. When the target starts to move, the position and speed of the target detection frame of the current frame are predicted from the target detection frame and target speed of the previous frame. This process describes the prediction approach of the Kalman filter. The update process of the Kalman filter is based on the predicted value of the previous frame and the observed value of the current frame (the predicted value and the observed value are in accordance with the normal distribution), which are linearly weighted to obtain the current prediction state of the system. The Hungarian algorithm then solves the matching

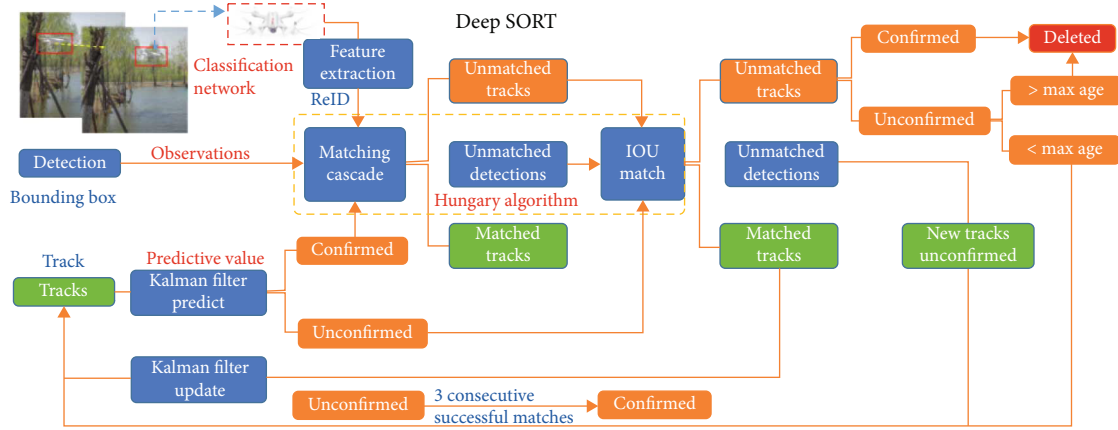


FIGURE 10: Deep SORT algorithm flow.

problem. After the similarity matrix, the Hungarian algorithm can solve the problem of matching the target of the two frames before and after the similarity matrix.

The most significant improvement of the Deep SORT algorithm over the SORT algorithm is the introduction of deep networks for appearance feature extraction and the use of models in pedestrian rerecognition for feature extraction. This operation also substantially reduces the amount of ID switching in the tracking algorithm. The Deep SORT algorithm can be divided into three steps: (i) predicting the trajectory tracks by Kalman filter; (ii) matching the predicted trajectory tracks with the detection frame in the current frame, including cascade matching and IOU matching, by using the Hungarian algorithm; and (iii) the third step which is to update the Kalman filter (Figure 10).

In matching the detection frame and the predicted trajectory, the situations shown in Figure 11 are expected to occur.

- (i) **Matched Tracks** in which the detection frame and the track match: ordinary continuous tracking targets are classified under this situation, and the targets in the previous and subsequent frames exist and can be matched.
- (ii) **Unmatched Detections** in which the detection box does not find a matching track: if the detector suddenly detects a new target in the image, the detection frame cannot find a matching target in the previous trajectory.
- (iii) **Unmatched Tracks** in which the track does not match the detection frame: the continuously tracked target disappears from the video or flies out of the shooting range of the camera, and the predicted trajectory does not find the matching detection frame information.
- (iv) Another situation occurring when two targets overlap: that is, when one target is occluded by another target, the trajectory of the occluded target cannot find a matching detection frame, and the target temporarily disappears from the image. When the occluded target reappears, the ID assigned by the



FIGURE 11: Four tracking situations.

occluded target should not change as much as possible and can be recognized by the algorithm as the target corresponding to the previous ID. This problem cannot be solved using the SORT algorithm. Thus, cascade matching in the Deep SORT algorithm is needed to solve it.

4. Experiments

4.1. Dataset. We generated a drone target dataset because of the lack of a publicly available one. To ensure the diversity of data sources and process the drone pictures downloaded from the Internet, we also prepared four drones for shooting: two quad-rotor drones and two single-rotor drones. We obtained drone flight shots in both indoor and outdoor scenarios. The video was shot at 30 frames per second to prevent the pictures from appearing too similar between the data. One of 10 video frames was selected as the dataset, and 2459 pictures were obtained. We also crawled 341 pictures from the Internet as a supplement. The dataset composition is listed in Table 1.

After the photos were acquired, the corresponding datasets were built for the YOLO and CenterNet networks. The YOLO dataset had its format requirements, and the CenterNet network used the COCO dataset format.

TABLE 1: UAV dataset.

Black four-rotor	White four-rotor	Yellow single rotor	Red single rotor	Total
823	678	500	799	2800

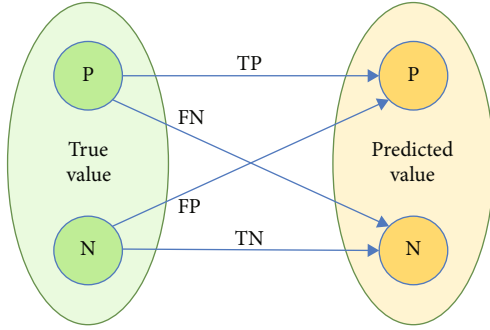


FIGURE 12: Four predicted results.

In the YOLO dataset format, a picture corresponds to a text annotation file, which contains the category and coordinate information of all target frames in the corresponding picture. The YOLO format requires that the coordinate values of all target frames be normalized. The conversion formula is as follows:

$$\begin{aligned}
 x' &= \frac{(x_{\min} + x_{\max})}{2W}, \\
 y' &= \frac{(y_{\min} + y_{\max})}{2H}, \\
 w' &= \frac{w}{W}, \\
 h' &= \frac{h}{H},
 \end{aligned} \tag{7}$$

where W and H are the width and height of the picture; x' and y' are the coordinates of the center point of the target frame, respectively; and w' and h' are the width and height of the target frame, respectively, and are normalized based on the width and height of the picture operation.

The COCO dataset format requires the coordinates of the upper left corner of the target frame and the width and height of the target frame. They are converted using the following formula:

$$\begin{aligned}
 x' &= x_{\min}, \\
 y' &= y_{\min}, \\
 w' &= x_{\max} - x_{\min}, \\
 h' &= y_{\max} - y_{\min},
 \end{aligned} \tag{8}$$

where x' and y' are the coordinates of the center point of the target frame and w' and h' are the width and height of the target frame, respectively.

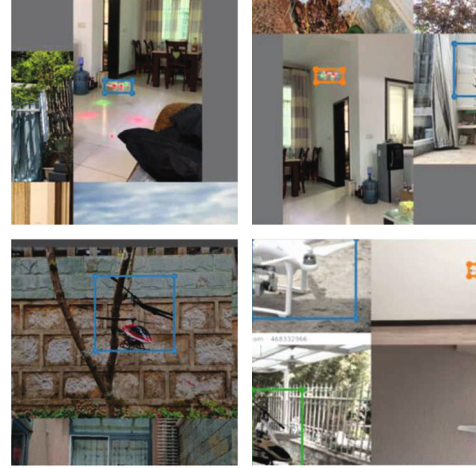


FIGURE 13: Image data splicing.

After obtaining the labeled data files, we divided the dataset into two parts with a train set:test set ratio of 9:1. The training set contained 2520 images, and the test set contained 280 images.

4.2. Experimental Performance Index

4.2.1. Target Detection Index

- (i) True positives (TP): the true value is a positive example, and the predicted value is a positive example.
- (ii) True negatives (TN): the true value is a negative example, and the predicted value is a negative example.
- (iii) False positives (FP): the true value is a negative example, and the predicted value is a positive example.
- (iv) False negatives (FN): the true value is a positive example, and the predicted value is a negative example.

As shown in Figure 12, the main indicators used are as follows:

- (i) Precision, the proportion of TP in the recognition result

$$\text{Precision} = \frac{\text{tp}}{\text{tp} + \text{fp}} \tag{9}$$

- (ii) Recall, where TP accounts for the proportion of all positive samples in the dataset

$$\text{Recall} = \frac{\text{tp}}{\text{tp} + \text{fn}} \tag{10}$$

- (iii) AP, where the size of the area is enclosed by the precision-recall curve
- (iv) Mean (mAP), the average of multiple categories of AP

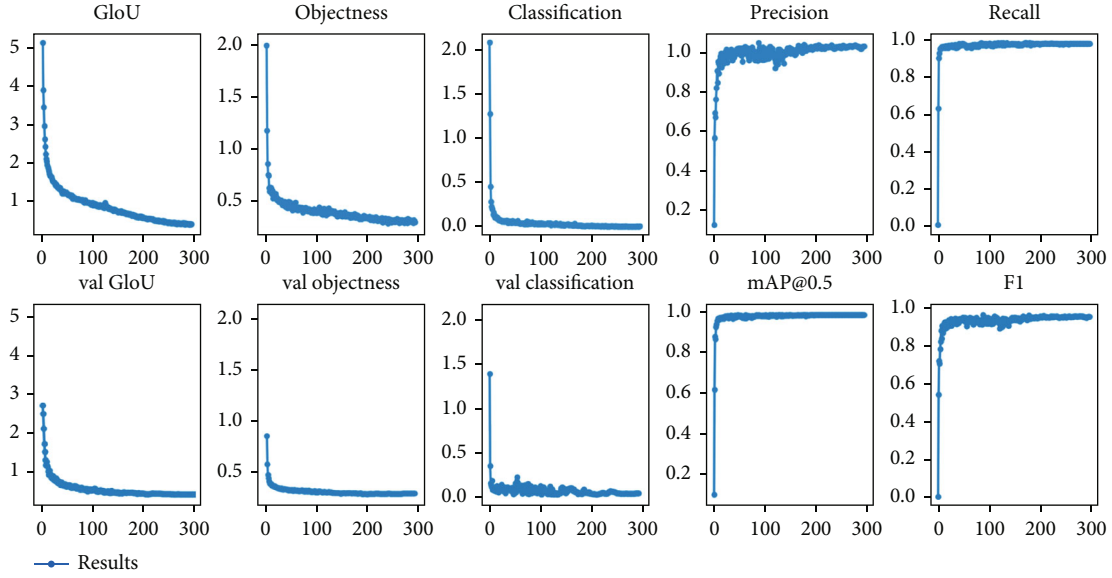


FIGURE 14: YOLOv3 loss curves.

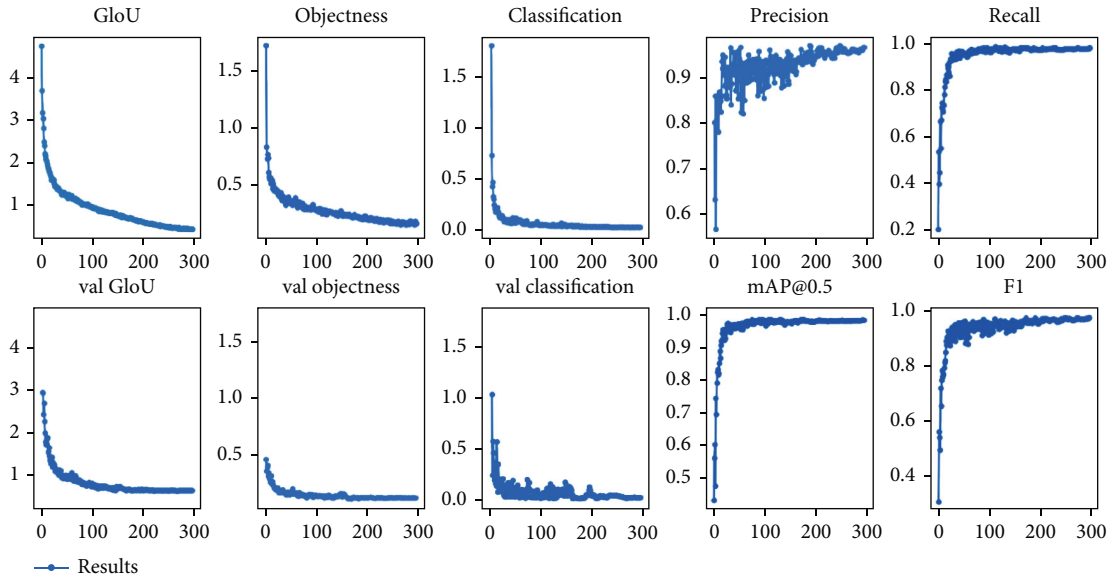


FIGURE 15: YOLOv3-SPP loss curves.

4.2.2. *Target Tracking Indicator.* To measure the target tracking algorithm performance, the following are the main indicators:

- (i) ID switch (IDSW), target ID switch total
- (ii) Fragmentation (FM), the total number of interrupted target tracking
- (iii) Multiple object tracking accuracy (MOTA):

$$MOTA = 1 - \frac{\sum_t (FN + FP + IDSW)}{\sum_t GT} \quad (11)$$

where GT is the ground true box of each frame.

4.2.3. *Training Result*

(1) *Detection Results.* YOLOv3, YOLOv3-SPP, and CenterNet algorithms were used to train the previously constructed drone dataset. The training machine environment was the i7-9700K CPU and the single card 1080 Ti GPU. The deep learning framework used by YOLOv3 was Pytorch1.4.0, and that by CenterNet was PyTorch 1.2.0. YOLOv3 and YOLOv3-SPP used the same training parameters. A batch size of 8300 epochs was trained, and CenterNet was trained for 150 epochs with a batch size of 16. By using this approach, the training volume of the three models was identical. Moreover, in the training process of YOLOv3, a data enhancement operation of randomly splicing data pictures was used (Figure 13), which could achieve multiscale training of the

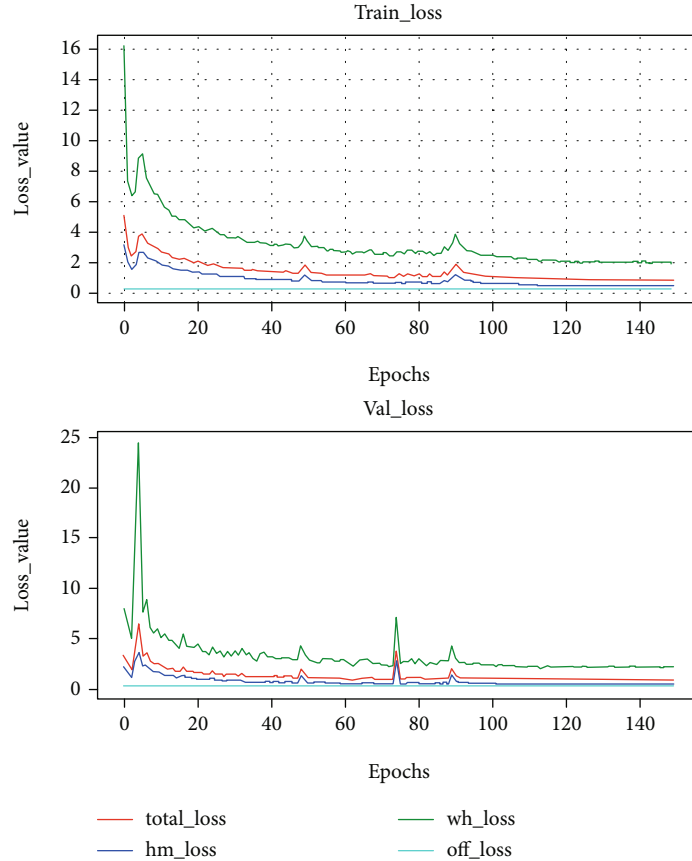


FIGURE 16: CenterNet loss curves.

TABLE 2: A comparison of model training time and detection performance.

Algorithm model	YOLOv3	YOLOv3-SPP	CenterNet-DLA34
Training time	12.5 h	9 h	8.5 h
mAP (IOU = 0.5)	0.988	0.993	0.958
Precision	0.971	0.982	0.961
Recall	0.962	0.973	0.942
Detection speed	69.5 FPS	69 FPS	23 FPS

target and also simulate the appearance of multiple targets in the same picture.

Figures 14–16 present the loss function curves during training of the three models:

All loss curves presented in Figure 16 meet the training expectations. Comparison and observation of the data in Table 2 reveal that the performances of YOLOv3 and YOLOv3-SPP do not considerably vary, although that of YOLOv3-SPP is slightly better. The speed of CenterNet can only reach one-third than that of YOLO, and the accuracy is lower than that of the YOLO model. Compared with YOLOv3, CenterNet eliminates the anchor setting and NMS operation calculation, and the final output feature map of CenterNet has only one heatmap layer. Therefore,

CenterNet should theoretically exhibit superior real-time performance than the YOLO series. However, CenterNet only needs one layer of feature maps; for such a layer to obtain sufficient feature information, a larger backbone network and a more complex feature fusion neck layer need to be used to extract features. Thus, the amount of calculation is not comparable to the YOLOv3 series. In summary, the single-stage detection model YOLOv3 under the anchor mechanism is evidently superior to CenterNet under the anchor-free mechanism on our UAV dataset.

(2) *Tracking Algorithm Experiment Results.* The core of Deep SORT matching tracking is to extract effective image feature information by using deep networks. The existing deep network is a suitable network model for pedestrian rerecognition. This model was originally a 751 classification network. The previous training data used 751 images of different positions and different angles as the training set for classification model training. The remaining problem was pedestrian reidentification.

In order to apply the algorithm to our drone information feature extraction, we used the detector to collect the drone image data, crop the image part in the detection frame, and classify and store it according to the categories of the four drones. We then used the previous pedestrian rerecognition model for classification training. The model obtained by

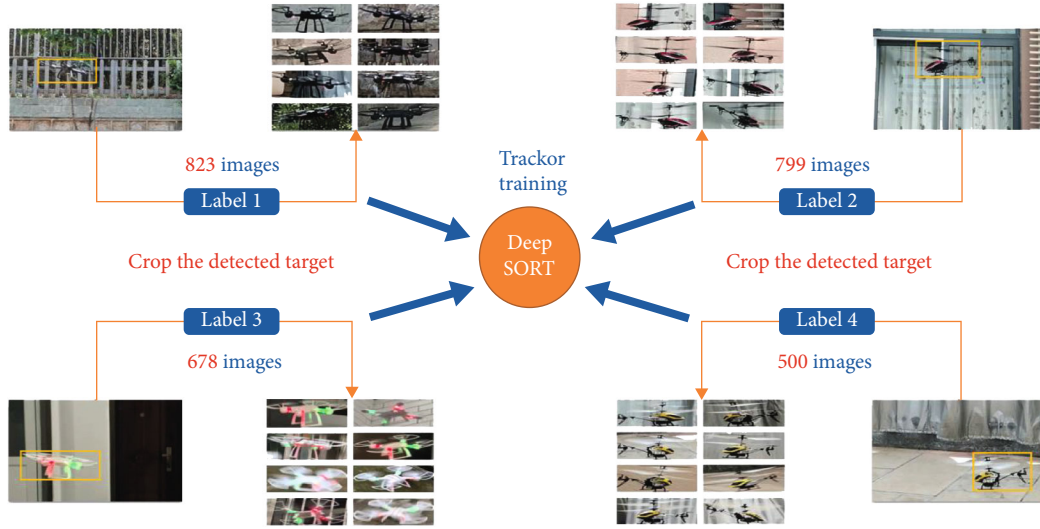


FIGURE 17: Generating data for Deep SORT.



FIGURE 18: Target tracking results in IoT.

TABLE 3: Target tracking performance.

Tracking model	YOLOv3-SPP + Deep SORT	CenterNet + Deep SORT
FP	0	0
FN	85	503
ID switch	4	31
FM	13	28
GT	1591	1591
MOTA	0.94406	0.66436
Speed (FPS)	54	25

classification training was employed as the feature extraction model of our tracking algorithm. The operation is illustrated in Figure 17.

Finally, the two detectors were combined with the tracker to obtain the following tracking results. The left picture in Figure 18 is the tracking effect of YOLOv3-spp + Deep SORT, and the right picture is the tracking effect of CenterNet + Deep SORT.

For the same online drone video tracking, the results are obtained as shown in Table 3.

5. Conclusion

In this study, we introduce a new method, referred to as tracking-by-detection, in tracking UAV targets. The

approach can achieve real-time high-precision tracking. This method used YOLOv3 as the detector and Deep SORT as the tracking mode to achieve a detection speed of 54 FPS and MOTA reaching 94.4%, which meets the requirements of real-time tracking of multiple targets for drones. On this basis, we also modify the YOLOv3 network by changing the loss function of the model in accordance with the characteristics of the drone target and adding the SPP module to collect the drone data to generate the initial anchor. These operations improve the MOTA of the modified YOLOv3-SPP network by 2% with high detection speed, which is only 5 FPS. For comparison, we also attempt tracking UAV targets by using the anchor-free mode. This mode is currently widely used in target tracking, with CenterNet as the target detector and Deep SORT as the tracker. The final detection speed of this model is 25 FPS, and the MOTA is 66.4%, only slightly enhancing real-time tracking. The experimental results indicate the effectiveness of the proposed approach and confirm that YOLOv3-SPP + Deep SORT is highly applicable for multitarget real-time tracking of UAVs.

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 declares that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (Project No. 2018YFB0505100). The work by Tao Hong is supported by the National Natural Science Foundation of China under Grant No. 61827901.

References

- [1] A. Fotouhi, M. Ding, and M. Hassan, "Understanding autonomous drone maneuverability for Internet of Things applications," in *2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, pp. 1–6, Macau, 2017.
- [2] F. Qi, X. Zhu, G. Mang, M. Kadoch, and W. Li, "UAV network and IoT in the sky for future smart cities," *IEEE Network*, vol. 33, no. 2, pp. 96–101, 2019.
- [3] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Mobile unmanned aerial vehicles (UAVs) for energy-efficient internet of things communications," *IEEE Transactions on Wireless Communications*, vol. 16, no. 11, pp. 7574–7589, 2017.
- [4] B. S. Oh, X. Guo, and Z. Lin, "A UAV classification system based on FMCW radar micro-Doppler signature analysis," *Expert Systems with Application*, vol. 132, pp. 239–255, 2019.
- [5] D. Han, D. Y. Gwak, and S. Lee, "Noise prediction of multirotor UAV by RPM fluctuation correction method," *Journal of Mechanical Science and Technology*, vol. 34, no. 4, pp. 1429–1443, 2020.
- [6] J. Wang, K. Chen, S. Yang, C. Loy, and D. Lin, "Region proposal by guided anchoring," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2965–2974, 2019.
- [7] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "Yolov4: optimal speed and accuracy of object detection," 2020, <https://arxiv.org/abs/2004.10934/>.
- [8] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," in *NIPS*, 2015.
- [9] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, 2016.
- [10] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," 2019, <https://arxiv.org/abs/1904.07850/>.
- [11] C. Kim, F. Li, A. Ciptadi, and J. M. Rehg, "Multiple hypothesis tracking revisited," in *IEEE International Conference on Computer Vision (ICCV)*, pp. 4696–4704, 2015.
- [12] P. Bergmann, T. Meinhardt, and L. Leal-Taixe, "Tracking without bells and whistles," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 941–951, 2019.
- [13] C. Feichtenhofer, A. Pinz, and A. Zisserman, "Detect to track and track to detect," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3057–3065, 2017.
- [14] S. H. Rezatofighi, A. Milan, Z. Zhang, Q. Shi, A. Dick, and I. Reid, "Joint probabilistic data association revisited," in *Proceedings of the IEEE international conference on computer vision*, pp. 3047–3055, 2015.
- [15] X. Zhou, V. Koltun, and P. Krähenbühl, "Tracking Objects as Points," in *European Conference on Computer Vision*, pp. 474–490, 2020.
- [16] Z. Dong, G. Li, Y. Liao, F. Wang, P. Ren, and C. Qian, "Centripetalnet: Pursuing high-quality keypoint pairs for object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10519–10528, 2020.
- [17] J. Shi, "Good features to track," in *1994 Proceedings of IEEE conference on computer vision and pattern recognition*, pp. 593–600, Seattle, WA, 1994.
- [18] C. Tomasi and T. Kanade, *Detection and tracking of point features. Technical Report CMUCS-91-132*, Carnegie Mellon University, 1991.
- [19] J. Redmon and A. Farhadi, "Yolov3: an incremental improvement," 2018, <http://arxiv.org/abs/1804.02767>.
- [20] Q. C. Mao, H. M. Sun, Y. B. Liu, and R. S. Jia, "Mini-YOLOv3: real-time object detector for embedded applications," *IEEE Access*, vol. 7, pp. 133529–133538, 2019.
- [21] R. A. Sturm, D. L. Duffy, Z. Z. Zhao et al., "A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color," *The American Journal of Human Genetics*, vol. 82, no. 2, pp. 424–431, 2008.
- [22] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*, pp. 448–456, 2015.
- [23] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1no. 14, pp. 281–297, 1967.
- [24] 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.
- [25] W. Liu, D. Anguelov, D. Erhan et al., "SSD: Single shot multi-box detector," in *European Conference on Computer Vision*, pp. 21–37, 2016.
- [26] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, 2017.
- [27] 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, 2016.
- [28] J. Dai, H. Qi, Y. Xiong et al., "Deformable convolutional networks," in *Proceedings of the IEEE international conference on computer vision*, pp. 764–773, 2017.
- [29] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *2017 IEEE international conference on image processing (ICIP)*, pp. 3645–3649, 2017.
- [30] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *2016 IEEE international conference on image processing (ICIP)*, pp. 3464–3468, 2016.

Research Article

Cooperative Offloading in D2D-Enabled Three-Tier MEC Networks for IoT

Jingyan Wu ^{1,2}, Jiawei Zhang ¹, Yuming Xiao ¹ and Yuefeng Ji ¹

¹State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China

²School of Information Engineering, Henan University of Science and Technology, Luoyang 471023, China

Correspondence should be addressed to Yuefeng Ji; jyf@bupt.edu.cn

Received 9 March 2021; Revised 20 May 2021; Accepted 19 July 2021; Published 16 August 2021

Academic Editor: Bo Rong

Copyright © 2021 Jingyan Wu 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.

Mobile/multi-access edge computing (MEC) takes advantage of its proximity to end-users, which greatly reduces the transmission delay of task offloading compared to mobile cloud computing (MCC). Offloading computing tasks to edge servers with a certain amount of computing ability can also reduce the computing delay. Meanwhile, device-to-device (D2D) cooperation can help to process small-scale delay-sensitive tasks to further decrease the delay of tasks. But where to offload the computing tasks is a critical issue. In this article, we integrate MEC and D2D cooperation techniques to optimize the offloading decisions and resource allocation problem in D2D-enabled three-tier MEC networks for Internet of Things (IoT). Mobile devices (MDs), edge clouds, and central cloud data center (DC) make up these three-tier MEC networks. They cooperate with each other to finish the offloading tasks. Each task can be processed by MD itself or its neighboring MDs at device tier, by edge servers at edge tier, or by remote cloud servers at cloud tier. Under the maximum energy cost constraints, we formulate the cooperative offloading problem into a mixed-integer nonlinear problem aiming to minimize the total delay of tasks. We utilize the alternating direction method of multipliers (ADMM) to speed up the computing process. The proposed scheme decomposes the complicated problem into 3 smaller subproblems, which are solved in a parallel fashion. Finally, we compare our proposal with D2D and MEC networks in simulations. Numerical results validate that the proposed D2D-enabled MEC networks for IoT can significantly enhance the computing abilities and reduce the total delay of tasks.

1. Introduction

Mobile devices (MDs) have limited computing resources and power capacity due to their portable sizes [1]. To cope with computation-intensive, delay-sensitive, and high-energy-cost tasks, mobile/multi-access edge computing (MEC) has been proposed to deploy resource-rich servers at base stations (BSs) within the proximity of MDs [2–5]. In Internet of Things (IoT), MEC has been advocated as a promising technique for providing massive MDs with enhanced computing and storage capabilities [6].

For one thing, due to the physical proximity, MEC can significantly reduce the transmission latency [7, 8] involving in communication compared with mobile cloud computing

(MCC) [9–12]. For another thing, MEC enables MDs to perform computation offloading, which can migrate their computing tasks [13] to resource-rich nodes and send results back to the MDs [14]. Offloading has always been a hot topic which can guarantee low computing time, as well as saving the battery energy of MDs [15], increasing the network throughput. Therefore, MEC can improve users' satisfaction [16] or quality-of-experience (QoE) of the end-users [17], relieve network congestion [18, 19], prolong the battery lifetime [20] of MDs, and further reduce the total latency distinctly. Thus, MEC is an efficient solution for IoT.

Nevertheless, offloading can obtain computing resources at the expense of extra transmission delay and associated energy consumption, due to the communication between

the MDs and edge servers over the wireless channels [21]. Offloading may be really beneficial to MDs with large amounts of computation and relatively small amounts of communication [22].

However, the edge servers associated with BSs have limited computing resources and can not always meet the requirements of the served MDs [23]. D2D cooperation can help to solve this problem, where the helpers (resource-rich MDs) can execute computing tasks from resource-limited MDs [24]. D2D-assisted cooperative offloading services play a complementary role in MEC, which is helpful when processing some small-scale tasks for energy-efficiency [25]. When a task has already been done well through D2D cooperation, accordingly, it can free up radio resources for other purposes and reduce the uplink delay of the task [26].

Therefore, in this paper, we propose cooperative computation offloading in D2D-enabled MEC networks for IoT aiming to minimize the total delay of tasks with the maximum energy consumption constraints. By leveraging the advantages of both MEC and D2D cooperation techniques, computing resources can be fully used. Our contributions can be summarized as follows:

- (1) We propose a computing architecture where cooperative offloading can be carried out in the D2D-enabled three-tier MEC networks. Each task can be processed locally by MD itself or its neighboring MDs at device tier, by edge servers at edge tier, or by remote cloud servers at cloud tier
- (2) The proposed three-tier MEC networks provide MDs with multiple optional offloading destinations. We formulate the offloading decisions and resource allocation in this MEC networks as an optimization problem aiming to minimize the total delay of tasks
- (3) The formulated problem is a mixed-integer nonlinear problem which is hard to solve. We decompose the problem into 3 subproblems. Then, we solve it utilizing the alternating direction method of multipliers (ADMM). Extensive simulation results validate that the proposed scheme is effective

The organization of this article is as follows. Section 2 provides a review of related work. Sections 3 presents the system model, including the network model, communication model, and computing model. In Section 4, we formulate the problem and give the equivalent form. In Section 5, we propose a parallel optimization framework to solve the problem and develop an efficient computation offloading scheme. Related simulations are provided in Section 6. Finally, we conclude this article in Section 7.

2. Related Work

Extensive research has been conducted on MEC. Many existing works paid attention to the computation offloading problem [15, 16, 20, 21, 27–32]. Some works jointly considered the computation offloading policy and the involved resource allocation [15, 33–35]. Others jointly considered the offload-

ing decisions, content caching [28, 36], and the resource allocation [20, 30]. All of these above works focused on two-tier edge computing networks.

Works focusing on the task offloading problem in three-tier edge computing networks can be found in [17, 37–40]. Tong et al. [39] first proposed a three-tier hierarchical edge cloud architecture to maximize the amount of peak mobile workloads from MDs being served. The performance of the proposed hierarchical edge cloud architecture was evaluated by a small-scale system experiment. The deployment of edge servers and cloud servers formed a three-tier MEC architecture, where local computing, edge computing, and cloud computing could coexist and cooperate to assist the task execution [38]. This scheme provided multiple offloading decisions for devices. The different offloading decisions may largely impact the network performance [21]. For example, offloading the tasks to the edge node or remote cloud DC will inevitably incur additional long communication delay [39], whereas executing the task locally may result in larger computing delay. Consequently, it is critical for devices to make proper offloading decisions in the three-tier computing networks.

Some works study cooperative offloading in MEC networks. du et al. [40] considered vertical cooperation between the fog and the cloud, in which the application can be processed in the MD locally, in the fog or in the cloud. The authors assumed that the offloading requests were usually very small; no buffer was needed for queueing the computing requests. Xiao and Krunz [17] designed optimal workload allocation solutions in a cooperative fog computing network. Instead of always relying on the cloud data center to process its unprocessed workload, each fog node can also forward part or all of its unprocessed workload to its neighboring fog nodes to further improve the QoE of its users. This paper solved the optimal workload allocation problem with the distributed ADMM. Wang et al. [38] developed a cooperative task offloading and computing resource allocation scheme in three-tier computing networks, considering the cooperation among the devices, the edge servers, and cloud servers vertically as well as the cooperation between the edge nodes horizontally. In these above works, they did not consider the cooperation among MDs.

D2D cooperation focuses on the collaborative computation among MDs. Feng et al. [24] developed a computation offloading scheme based on D2D communications, in which resource-limited MDs could offload their computation-intensive tasks to appropriate nearby resource-rich MDs. Xing et al. [41] studied D2D-enabled multihelper MEC networks, where a local user could be helped by its nearby wireless devices serving as helpers for cooperative computation. Literature [26] considered offloading computational tasks to nearby devices or to an edge cloud and developed a decentralized algorithm to allocate the computational tasks among them. However, these works were not carried out in three-tier networks.

In MEC networks, there are horizontal collaboration at the device tier [41] and the edge tier [17] as well as the vertical collaboration [40] among MDs, edge nodes, and cloud nodes [2]. Different from the above literatures, our paper

proposes cooperative computation offloading in D2D-enabled three-tier MEC networks for IoT. Each task can be processed locally by MD itself or its neighboring MDs, by edge servers at edge tier, or by cloud servers at cloud tier.

3. System Model

In this section, the system model adopted in this paper is described. We introduce the proposed network model and present the communication and the computing model in detail. Finally, we calculate the total delay of tasks and energy consumption.

3.1. Network Model. As depicted in Figure 1, we propose a D2D-enabled three-tier MEC network architecture as a part of the IoT, which consists of N MDs labeled as $\mathcal{N} = \{1, 2, \dots, N\}$, M base stations (BSs), and a remote central cloud DC equipped with many remote cloud servers. Each MD has the ability of D2D communication and is connected to the closest BS via a wireless link. Each BS is equipped with some edge servers in edge cloud at edge tier and can be regarded as an edge node. The BSs are connected to the remote cloud servers in central cloud DC at cloud tier via a low-latency wired backhaul link (e.g., optical link) [4]. Optical backhauling is a promising solution [42, 43] to offer high throughput, low latency, and reduced energy consumption. Moreover, the flexibility and reliability can be effectively improved by the adoption of artificial intelligence-driven autonomous optical networks [44].

In this D2D-enabled cooperative architecture, we assume that each MD has a computing task to be processed locally by MD itself or its neighboring MDs at device tier, by edge cloud, or by central cloud DC. MDs can help each other when they are free or have surplus computing resources. The helper may be rewarded in terms of accumulated reward points. MDs with more accumulated reward points can be helped with higher precedence when they need help. Multiple neighboring local MDs can communicate with each other through WiFi [45] or Bluetooth [46]. The network operator (e.g., BS) can gather sufficient D2D connectivity information through network-assisted device discovery and local information reporting by the devices [25]. If the network condition is bad, the task of each local MD is small and/or delay-sensitive; MDs may be more willing to choose D2D cooperation due to short transmission time. That is to say, these tasks prefer to be processed among local MDs.

The computation offloading process is shown in Figure 2 [35]. Each MD can send an offloading request including the information of the MD such as its local processing capability and power consumption, the properties of the task such as the size and emergency of the task, and the maximum tolerable delay [40]. According to this collected information, MDs finally decide where should the tasks be processed. The computing task of MD n can be completed locally within its CPU, or by its neighboring MD via D2D cooperation, or remotely in the edge cloud via MEC offloading, or further up the cloud DC via cloud offloading [35].

After being processed in one of these nodes, the result of the computing task is transmitted back to the corresponding

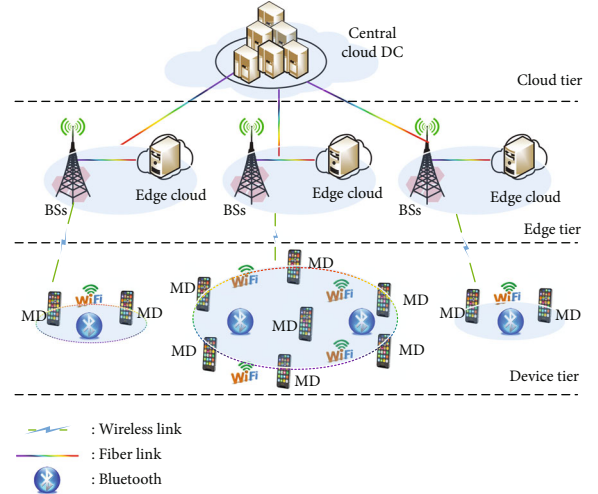


FIGURE 1: A D2D-enabled three-tier MEC network.

MD. Similar to many previous works about MEC [18, 20, 21, 26], we neglect the return result time of the computed tasks in this work. The assumption is justified for many applications including face recognition and anomaly detection, where the size of the result is much smaller than the size of the uplink data. We do not consider the queueing of offloading tasks, as in [19, 40]. To obtain tractable analysis, we also omit the delay due to decision-making [20].

3.2. Communication Model. For the communication model, the computing task of MD n is described by $\Gamma_n = (L_n, \tau_n^{\max}, C_n)$, where L_n (in bits) stands for the input data size of task Γ_n , τ_n^{\max} (in s) stands for the task computation deadline for MD n , and C_n (in CPU cycles) stands for the number of required CPU cycles to accomplish the task Γ_n .

When a task of an MD can not be processed well by MD itself or its neighboring MDs, the task can be sent to the BS. The bandwidth resource of the BS is denoted as B Hz. However, the limited radio bandwidth needs to be allocated among all the edge-processing and the cloud-executing MDs for communication with the edge node. Within our communication model, all the devices equally share the bandwidth. The uplink bandwidth of MD n is B_n Hz; the uplink achievable transmission rate R_n for radio access can be approximated by Shannon's formula:

$$R_n = B_n \cdot \log_2 \left(1 + \frac{p_n B_n G_n}{\sigma^2} \right), \quad (1)$$

where p_n is the transmission power density of MD n , G_n stands for the channel gain between the MD n and the BS, and σ^2 denotes the power of additive white Gaussian noise.

Furthermore, the uplink transmission delay for MD n offloading the task Γ_n to the BS is calculated as

$$T_n^{tr} = \frac{L_n}{R_n}. \quad (2)$$

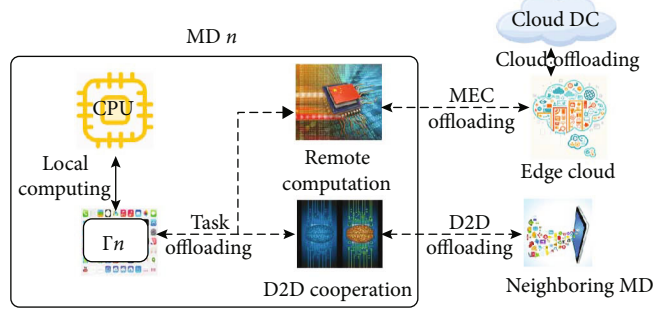


FIGURE 2: Computation offloading process.

And the transmission energy consumption for MD n offloading the task Γ_n to the BS can be given as

$$E_n^{tr} = P_n T_n^{tr}, \quad (3)$$

where P_n is the total power consumption of MD n for transmission in the uplink, consisting of the static power consumption and transmit power [38].

3.3. Computing Model. We denote the offloading decisions for MD n by $x_{n,j}, y_n, z_n \in \{0, 1\}$, where $x_{n,j} = 1, y_n = 1, z_n = 1$ indicate that the task of MD n is executed by MD itself (e.g., $n = j$) or its neighboring MD j , by edge server, or by cloud server, respectively; otherwise, $x_{n,j} = 0, y_n = 0, z_n = 0$. Thus, the offloading decisions of device n is constrained by

$$\sum_{j=1}^N x_{n,j} + y_n + z_n = 1, \quad \forall n \in \mathcal{N}. \quad (4)$$

In the constraint, it should be noted that one and only one of the offloading decisions $x_{n,j}, y_n,$ and z_n for the MD n can be 1 at any time. It also implies that the task of each MD is impartible in this paper.

Next, we will discuss the computation overhead in terms of processing delay and energy consumption of MD n with different offloading approaches.

3.3.1. Local Processing. For the local task processing approach, let f_n^L denote the local computing capability (in CPU cycles/s) of MD n . Then, the computing delay of processing the task Γ_n locally can be

$$T_n^L = T_{\Gamma_n}^n = \frac{C_n}{f_n^L}. \quad (5)$$

The CPU power consumption is a function of execution frequency f_n^L . We can compute the energy consumption of MD n as [47, 48]

$$E_n^L = 10^{-11} C_n (f_n^L)^2. \quad (6)$$

From (5) and (6), we can notice that both T_n^L and E_n^L are only related to C_n and f_n^L , which are the known inherent features of MD n and its computing task.

3.3.2. D2D Cooperation. For D2D cooperation, we define $d_{n,j}$ as the D2D link between local MD n and j , P_n^D as the power consumption of MD n , and $T_{n,j}^{tr}$ as the round trip time (RTT) for task transferring between the D2D pairs. This RTT is assumed to be stable during an offloading period, while it may change across different periods. It can be estimated at the beginning of the offloading period, by using the history information or empirical formula [49]. Thus, when the task Γ_n is executed finally at MD j , the total delay T_n^D consists of the communication delay $T_{n,j}^{tr}$ between MD n and MD j and the processing delay $T_{\Gamma_n}^j$ at MD j . It can be calculated as

$$T_n^D = T_{\Gamma_n}^j + T_{n,j}^{tr}, \quad \forall j \in \mathcal{N}, \quad (7)$$

according to (5), $T_{\Gamma_n}^j = C_n / f_j^L$. Note that $T_{n,j}^{tr} = 0$ when $n = j$, because there is no task to transfer within the local MD. Thus, the delays of tasks in the device tier all meet (7).

In this case, the energy cost of MD n is mainly the transmission energy consumption. It can be given as

$$E_n^D = P_n^D T_{n,j}^{tr}. \quad (8)$$

3.3.3. Edge Computing. For the edge computing approach, the total computing capacity of the BS is denoted as F (in CPU cycles/s), which can be allocated to the computing tasks processed by the BS. Denote f_n^B as the computing resource allocated to MD n by the BS. Thus, the computing delay of MD n executed by BS is given as

$$T_n^B = \frac{C_n}{f_n^B}. \quad (9)$$

When the tasks are computed in the edge server, the MD n sends the task Γ_n to the associated BS. Then, the BS will process the offloaded task by employing its computing

resources. Combining with equation (2), the total delay of task Γ_n processed by the BS is calculated by

$$T_n^E = T_n^{tr} + T_n^B. \quad (10)$$

3.3.4. Cloud Computing. For the computing task offloaded to the cloud DC, MD n firstly transmits the task Γ_n to the nearest BS through a wireless link. Then, the BS forwards the task Γ_n to the cloud DC through a fiber optical link. Since the cloud servers have plenty of computing resources and the fiber optical link between the edge cloud and the cloud DC is of sufficiently large capacity, the allocation of these resources will not be discussed here. Ordinarily, the computing capability of the cloud DC is much higher than that of the MEC servers.

Let the RTT for task transmission between the edge server at BS and cloud server be $T_{B,C}$. Denote f_n^C as the cloud computing capability (in CPU cycles/s) assigned to MD n , which is assumed to be a predetermined value according to the cloud computing service [40]. Therefore, the cloud computing delay of MD n is given by

$$T_n^{fC} = \frac{C_n}{f_n^C}. \quad (11)$$

Then, combining with equation (2), the total delay of cloud processing for MD n can be expressed as

$$T_n^C = T_n^{tr} + T_{B,C} + T_n^{fC}. \quad (12)$$

The above notations that will be used in this paper are summarized in Table 1.

As mentioned earlier, we ignore the return result time of the computed tasks. According to (2)–(12), the total task delay and energy consumption of MD n are calculated, respectively, as

$$T_n = \sum_{j=1}^N x_{n,j} T_n^D + y_n T_n^E + z_n T_n^C, \quad (13)$$

$$E_n = x_{n,n} E_n^L + \sum_{j \in \mathcal{N} \setminus \{n\}} x_{n,j} E_n^D + \left(1 - \sum_{j=1}^N x_{n,j}\right) E_n^{tr}. \quad (14)$$

4. Problem Formulation and Transformation

4.1. Problem Formulation. To minimize the total delay, we formulate the task offloading and resource allocation problem with the maximum energy cost constraints. Let the offloading strategy vector of MD n as $\Theta_n = \{x_{n,1}, x_{n,2}, \dots, x_{n,N}, y_n, z_n\}$ and the offloading decision profile of all devices as $\Theta = \{\Theta_n, n \in \mathcal{N}\}$. The computing resource allocation vector of all devices is denoted as $\Omega = \{f_n^B, n \in \mathcal{N}\}$. The maximum energy cost of MD n is denoted as E_n^{\max} (e.g., the

battery capacity). As a result, the joint task offloading and resource allocation problem are formulated as

$$P1 : \min_{\Theta, \Omega} \sum_{n=1}^N T_n \quad (15)$$

$$\text{s.t.} \quad \sum_{j=1}^N x_{n,j} + y_n + z_n = 1, \quad \forall n \in \mathcal{N}, \quad (16)$$

$$x_{n,j}, y_n, z_n \in \{0, 1\}, \quad \forall n \in \mathcal{N}, \quad (17)$$

$$\sum_{n=1}^N f_n^B \leq F, \quad (18)$$

$$x_{n,n} E_n^L + \sum_{j \in \mathcal{N} \setminus \{n\}} x_{n,j} E_n^D + \left(1 - \sum_{j=1}^N x_{n,j}\right) E_n^{tr} \leq E_n^{\max}. \quad (19)$$

The constraints (16) and (17) denote that each task is processed locally by MD itself or by its neighboring MDs, by edge servers, or by cloud servers. The constraint (18) states the computing resource consumption of all the MDs can not exceed the computing capacity of the edge servers. The constraint (19) indicates the maximum energy cost of MD n . Based on (7), (10), (12), and (13), the total delay of MD n can be written as

$$\begin{aligned} T_n &= \sum_{j=1}^N x_{n,j} T_n^D + y_n T_n^E + z_n T_n^C \\ &= \sum_{j=1}^N x_{n,j} \left(T_{\Gamma_n}^j + T_{n,j}^{tr} \right) + y_n \left(T_n^{tr} + T_n^B \right) \\ &\quad + z_n \left(T_n^{tr} + T_{B,C} + T_n^{fC} \right). \end{aligned} \quad (20)$$

The optimization problem $P1$ is a mixed-integer nonlinear problem and is difficult to solve due to the following observation. Since that $x_{n,j}$, y_n , and z_n are binary variables, the feasible set of problem (15) is not convex, so that the objective function of problem (15) is not convex. As is shown, problem (15) is a mixed discrete and nonconvex optimization problem, and this kind of problem is usually considered as NP-hard [20]. Then, we rewrite the objective function (15) as

$$\begin{aligned} \Psi(\Theta) &= \sum_{n=1}^N T_n = \sum_{n=1}^N \sum_{j=1}^N x_{n,j} \left(T_{\Gamma_n}^j + T_{n,j}^{tr} \right) + \sum_{n=1}^N y_n \left(T_n^{tr} + T_n^B \right) \\ &\quad + \sum_{n=1}^N z_n \left(T_n^{tr} + T_{B,C} + T_n^{fC} \right). \end{aligned} \quad (21)$$

In that, $P1$ is a convex problem with all binary variables fixed. We can use the branch-and-bound method to find a near-optimal solution [38]. But as far as our objective

TABLE 1: Notation and corresponding definition.

Notation	Definition
Γ_n	Computing task of MD n
L_n	Data size of task Γ_n (in bits)
C_n	CPU cycles required to accomplish the task Γ_n
B	The bandwidth of the BS
B_n	The uplink bandwidth of MD n
R_n	The uplink transmission rate of MD n
p_n	The transmission power density of MD n
P_n^D	The power consumption of MD n in D2D cooperation
P_n	The total power consumption of MD n
G_n	The channel gain between the MD n and the BS
σ^2	The power of the additive white Gaussian noise
f_n^L	The local computing capability of MD n (in CPU cycles/s)
f_n^B	The computing resource allocated to MD n by the BS
f_n^C	The computing resource allocated to MD n by the cloud
F	The computing capacity of the BS
τ_d^{\max}	The task computation deadline for MD n
T_n^{tr}	The uplink transmission delay of task Γ_n offloaded to the BS
$T_{\Gamma_n}^j$	The computing delay of task Γ_n processed at MD j
$T_{n,j}^{tr}$	The average round trip time between MD n and j
T_n^L	The delay of task Γ_n processed locally
T_n^D	The total delay of task Γ_n via D2D cooperation
T_n^B	The computing delay of MD n by the BS
$T_{B,C}$	The round trip time between BS and the cloud servers
T_n^{fC}	The cloud computing delay of MD n
T_n^C	The total delay of task Γ_n processed by the cloud servers
E_n^{tr}	The transmission energy consumption of task Γ_n to the BS
E_n^L	The energy consumption of processing the task Γ_n locally
E_n^D	The energy consumption of processing Γ_n via D2D cooperation

function (15) is concerned, we should note that the complexity of the method remains exponential. The scale of the problem becomes large as the number of MDs grows, which makes it hard to obtain the solution in a reasonable amount of time. Then, in the next section, we will decouple the optimization problem (15) to present an effective computation offloading scheme ADMM to solve the problem $P1$.

4.2. Problem Transformation. To make it possible for each offloading decisions participating in the computation, we will decompose the problem $P1$ into several small subproblems, so that it can be solved in a distributed manner. Thus, we need to properly handle the coupling constraint among variables. Specially speaking, the optimization variables $x_{n,j}$, y_n , and z_n are coupled in the constraints (16), which make the problem inseparable. Thus, to make the problem separable,

we introduce the local copies of the global variables and denote them as $\hat{x}_{n,j}$, \hat{y}_n , and \hat{z}_n , respectively. Then, we have

$$\begin{cases} \hat{x}_{n,j} = x_{n,j}, & \forall n \in \mathcal{N}, j \in \mathcal{N}, \\ \hat{y}_n = y_n, & \forall n \in \mathcal{N}, \\ \hat{z}_n = z_n, & \forall n \in \mathcal{N}. \end{cases} \quad (22)$$

Based on equation (22), the coupling constraint of (16) can be written as

$$\sum_{j=1}^N \hat{x}_{n,j} + \hat{y}_n + \hat{z}_n = 1, \quad \forall n \in \mathcal{N}. \quad (23)$$

Let $\hat{\Theta} = \{\hat{x}_{n,1}, \hat{x}_{n,2}, \dots, \hat{x}_{n,N}, \hat{y}_n, \hat{z}_n, n \in \mathcal{N}\}$ be the local variables vector of all MDs. Then, we give the equivalent global consensus version of $P1$ by substituting (16) with (22) and (23) shown as

$$\begin{aligned} P2 : \min_{\Theta, \Omega, \hat{\Theta}} \quad & \Psi(\Theta) \\ \text{s.t.} \quad & (17), (18), (19), (22), (23). \end{aligned} \quad (24)$$

Now, it is obvious that in problem $P2$, the objective functions $\Psi(\Theta)$ with feasible sets are separable. But the consensus constraints (23) remain coupled. As far as the problem $P2$ is concerned, the numbers of variables and constraints reach $N^2 + 2N$ and $2N^2 + 4N$, respectively. The size of this problem becomes extremely large as the number of MDs grows.

ADMM is a promising approach to solve the large-scale optimization problem [20], which is originally devised to solve convex optimization problems. Recently, it has been further explored to solve nonconvex problems due to the flexibility of the ADMM framework [30]. In the next section, we will apply ADMM to solve the problem in a distributed fashion.

5. Efficient Computation Offloading Scheme

In this section, we develop a parallel optimization framework based on ADMM and difference of convex function (D.C.) programming. First, we derive the augmented Lagrangian function. With corresponding global consensus constraints, we formulate the ADMM iteration steps [28, 38, 50]. Then, the update methods for ADMM iterations are presented, and the overall algorithm is summarized.

5.1. Augmented Lagrangian and ADMM Variables. According to [50], the augmented Lagrangian function is expressed as

$$\begin{aligned}
L_\rho(\Theta, \widehat{\Theta}, \lambda, \mu, \gamma) & \\
&= \varphi(\Theta) + \sum_{n=1}^N \sum_{j=1}^N \lambda_{n,j} (x_{n,j} - \widehat{x}_{n,j}) + \sum_{n=1}^N \mu_n (y_n - \widehat{y}_n) \\
&+ \sum_{n=1}^N \gamma_n (z_n - \widehat{z}_n) + \frac{\rho}{2} \sum_{n=1}^N \sum_{j=1}^N (x_{n,j} - \widehat{x}_{n,j})^2 \\
&+ \frac{\rho}{2} \sum_{n=1}^N (y_n - \widehat{y}_n)^2 + \frac{\rho}{2} \sum_{n=1}^N (z_n - \widehat{z}_n)^2,
\end{aligned} \tag{25}$$

where $\lambda = \{\lambda_{n,j}\}_{n,j \in \mathcal{N}}$, $\mu = \{\mu_n\}_{n \in \mathcal{N}}$, and $\gamma = \{\gamma_n\}_{n \in \mathcal{N}}$ are the Lagrange multipliers with respect to the constraint (23), respectively. $\rho > 0$ is the so-called penalty parameter, augmented Lagrangian parameter, which is a constant parameter intend to adjust the convergence speed of ADMM [50].

With the ADMM method to solve the problem $P2$, the resulting optimization steps are described as follows.

5.2. Global Variable Update. The global variables $\{\Theta, \lambda, \mu, \gamma\}$ are updated in the $[t + 1]$ -th iteration by settling the following optimization problem [51]:

$$\begin{aligned}
P3 : \min_{\Theta, \Omega} L_\rho(\Theta, \widehat{\Theta}^{[t]}, \lambda^{[t]}, \mu^{[t]}, \gamma^{[t]}) \\
\text{s.t. } (17), (18), (19),
\end{aligned} \tag{26}$$

where the superscript $[t]$ stands for the iteration index.

We denote $x_n = \{x_{n,j}, j \in \mathcal{N}\}$ as the local offloading decision vector of MD n , $y = \{y_n, n \in \mathcal{N}\}$ as the decision vector of offloading tasks to BS j . For cloud computing, $z = \{z_n, n \in \mathcal{N}\}$ is the decision vector of offloading tasks to the cloud DC. Then, omitting the local copies of the variables $\widehat{x}_{n,j}$, \widehat{y}_n , \widehat{z}_n , the objective function in $P3$ can be rewritten as

$$L_\rho(\Theta, \widehat{\Theta}^{[t]}, \lambda^{[t]}, \mu^{[t]}, \gamma^{[t]}) = \sum_{j=1}^M f_j(x_n) + g(y) + h(z), \tag{27}$$

where the functions $f_j(x_n)$, $g(y)$, and $h(z)$ are given

$$f_j(x_n) = \sum_{n \in \mathcal{N}} \left[\frac{\rho}{2} (x_{n,j} - \widehat{x}_{n,j}^{[t]})^2 + (T_n^L + \lambda_n^{[t]}) x_{n,j} \right], \tag{28}$$

$$g(y) = \sum_{n \in \mathcal{N}} \left[\frac{\rho}{2} (y_n - \widehat{y}_n^{[t]})^2 + (T_n^E + \mu_n^{[t]}) y_n \right], \tag{29}$$

$$h(z) = \sum_{n \in \mathcal{N}} \left[\frac{\rho}{2} (z_n - \widehat{z}_n^{[t]})^2 + (T_n^C + \gamma_n^{[t]}) z_n \right]. \tag{30}$$

Therefore, it is easy to recognize that the objective function and the feasible region in $P3$ are separable completely. As a result, $P3$ can be decomposed equivalently into the following three subproblems ($P1'$, $P2'$, and $P3'$), which are given by

$$\begin{aligned}
P1' : \min_{x_n} f_j(x_n) \\
\text{s.t. } (18), (19), x_{n,j} \in \{0, 1\}, \quad \forall n \in \mathcal{N},
\end{aligned} \tag{31}$$

$$\begin{aligned}
P2' : \min_y g(y) \\
\text{s.t. } y_n \in \{0, 1\}, \quad \forall n \in \mathcal{N}.
\end{aligned} \tag{32}$$

Meanwhile, the subproblem $P3'$ related to the cloud servers is written as

$$\begin{aligned}
P3' : \min_z h(z) \\
\text{s.t. } z_n \in \{0, 1\}, \quad \forall n \in \mathcal{N}.
\end{aligned} \tag{33}$$

5.2.1. The Solution of the First Subproblem $P1'$. There are some difficulties in solving this subproblem $P1'$ with the existence of the binary variable constraint. To overcome the difficulties, an equivalent transformation of the original problem is necessary. We can transform the binary constraints in $P1'$ as

$$\sum_{n \in \mathcal{N}} (x_{n,j} - x_{n,j}^2) \leq 0, \tag{34}$$

$$0 \leq x_{n,j} \leq 1, \quad \forall n \in \mathcal{N}. \tag{35}$$

By replacing the binary constraint with (34) and (35), problem $P1'$ is rewritten as

$$\begin{aligned}
P4 : \min_{x_n} f_j(x_n) \\
\text{s.t. } (18), (19), (34), (35), \quad \forall n \in \mathcal{N}.
\end{aligned} \tag{36}$$

After the above equivalent transformations, the objective function of $P4$ is linear and constraints except (34) are convex. To deal with (34), we further transform $P4$ as given by

$$P5 : \min_{x_j} f_j(x_j) + \alpha \sum_{n \in \mathcal{N}} (x_{n,j} - x_{n,j}^2) \tag{37}$$

$$\text{s.t. } (18), (19), (35), \quad \forall n \in \mathcal{N},$$

where α acts as a penalty factor. It is proven for a sufficiently large amount; the problem $P5$ is equivalent to $P4$ [52]. Here, we define $\varphi_j(x_n)$ and $\phi_j(x_{n,j})$ as

$$\varphi_j(x_n) = f_j(x_n) + \alpha \sum_{n \in \mathcal{N}} x_{n,j}, \tag{38}$$

$$\phi_j(x_{n,j}) = \alpha \sum_{n \in \mathcal{N}} x_{n,j}^2. \tag{39}$$

Then, it can be found that the objective function can be written as the difference of the two convex functions $\varphi_j(x_n)$ and $\phi_j(x_{n,j})$. We can find that $P5$ is a D.C. programming problem [53]. Therefore, to solve $P5$, we can exploit the sequential convex approximation [24] of $\phi_j(x_{n,j})$ in the x_n

domain. $\phi_j(x_{n,j})$ can be approximated by its first-order Taylor expansion, and the objective function can be derived as

$$\varphi_j(x_n) - \phi_j(x_{n,j}) + \nabla_{x_j} \phi_j(x_{n,j}) \times (x_{n,j} - x_{n,j}^{([t])}), \quad (40)$$

where $\nabla_{x_n} \phi_j(x_{n,j})$ is the gradient of $\phi_j(x_{n,j})$ at x_n and can be expressed as $\nabla_{x_n} \phi_j(x_{n,j}) = (\partial \phi_j(x_{n,j})) / \partial x_{n,j}, \forall j \in \mathcal{N}$.

Then, we can obtain an approximation solution to get a local optimal solution $x_n^{[t]}$ during each iteration, which is sketched in Algorithm 1, where $x_n^{[t]}$ is the solution at the t -th iteration. As a result, P5 is a convex problem and can be solved efficiently by the standard convex optimization methods [54]. We can obtain near-optimal solution $x_n^{[t]}$ of P5 by iteratively solving.

It is worth noting that the solution $x_n^{[t]}$ is better than the previous solution $x_n^{[t-1]}$. Since the constraint set is compact, the sequence $\{x_n^{[t]}\}_{t=1,2,3,\dots,n}$ is proved to be convergent [52] by Cauchy theorem. The iterative process can stop after finite iterations. Given an initial point, due to the convexity of each iteration, Algorithm 1 converges to the same solution. In the issue, the convergence properties in Algorithm 1 could facilitate the convergence of ADMM.

5.2.2. The Solutions of the Subproblem P2'. For the subproblem P2', it is an unconstrained binary optimization problem. Due to the fact that possible values of y_n are 0 and 1, we have $y_n^2 = y_n$. Then, the expression after the summation sign of (29) is

$$\frac{\rho}{2} (y_n - 2y_n \hat{y}_n^{[t]} + \hat{y}_n^{[t]2}) + (T_n^E + \mu_n^{[t]}) y_n. \quad (41)$$

By comparing the corresponding objective function values of two possible solutions, the solution can be given by

$$y_n = \begin{cases} 0 & \frac{\rho}{2} - \rho \hat{y}_n^{[t]} + T_n^E + \mu_n^{[t]} > 0, \\ 1 & \frac{\rho}{2} - \rho \hat{y}_n^{[t]} + T_n^E + \mu_n^{[t]} \leq 0. \end{cases} \quad (42)$$

5.2.3. The Solutions of the Subproblem P3'. For the subproblem P3', it is also an unconstrained binary optimization problem as P2'. Thus, the solution can also be given as

$$z_n = \begin{cases} 0 & \frac{\rho}{2} - \rho \hat{z}_n^{[t]} + T_n^C + \gamma_n^{[t]} > 0, \\ 1 & \frac{\rho}{2} - \rho \hat{z}_n^{[t]} + T_n^C + \gamma_n^{[t]} \leq 0. \end{cases} \quad (43)$$

5.3. Local Variable Update. Now, we move on to the local variables. After obtaining the global variables, at the $[t+1]$ -th iteration, the local variables $\hat{\Theta}$ are updated by solving the following optimization problem:

1. Initialization:

Choose an initial feasible solution $\{x_{n,j}^{(0)}\}$ of P1', $t=0$.

2. Iteration:

repeat

Solve the following convex problem P5:

$$\min_{x_n} \varphi_j(x_n) - \phi_j(x_{n,j}) + \nabla_{x_n} \phi_j(x_{n,j}) \times (x_{n,j} - x_{n,j}^{([t])})$$

s.t. (18),(19),(35)

obtain the optimal solution $x_n^{[t+1]}$

update $t=t+1$

3. until convergence of x_n .

ALGORITHM 1: D.C.-based algorithm for subproblem P1'.

$$P6 : \min_{\hat{\Theta}} L_{\rho} \left(\Theta^{[t+1]}, \hat{\Theta}^{[t]}, \lambda^{[t]}, \mu^{[t]}, \gamma^{[t]} \right) \quad (44)$$

$$\text{s.t. (23), } \hat{x}_{n,j}, \hat{y}_n, \hat{z}_n \in \{0, 1\}, \quad \forall n \in \mathcal{N}, j \in \mathcal{N}.$$

After eliminating some constant terms from the objective function, it is equivalent to solve the following problem:

$$P7 : \min_{\hat{\Theta}} \sum_{n=1}^N F(\hat{\Theta}_n)$$

$$\text{s.t. (23), } \hat{x}_{n,j}, \hat{y}_n, \hat{z}_n \in \{0, 1\}, \quad \forall n \in \mathcal{N}, j \in \mathcal{N}, \quad (45)$$

where the function $F(\hat{\Theta}_n)$ is given by

$$F(\hat{\Theta}_n) = \sum_{j=1}^M \left[\frac{\rho}{2} \hat{x}_{n,j}^2 - \left(\rho x_{n,j}^{[t+1]} + \lambda_{n,j}^{[t]} \right) \hat{x}_{n,j} \right]$$

$$+ \left[\frac{\rho}{2} \hat{y}_n^2 - \left(\rho y_n^{[t+1]} + \mu_n^{[t]} \right) \hat{y}_n \right]$$

$$+ \left[\frac{\rho}{2} \hat{z}_n^2 - \left(\rho z_n^{[t+1]} + \gamma_n^{[t]} \right) \hat{z}_n \right]. \quad (46)$$

As mentioned earlier, the constraint (23) implies that only one of $x_{n,j}$, y_n , and z_n for the MD n can be 1 at any time. Thus, there are total $N+2$ possible solutions for each MD n . We only need to calculate the corresponding objective function values of the $N+2$ different offloading decisions and choose the minimum one, which corresponds to the optimal solution. Formally, the optimal solution of P6 is written as

$$\begin{cases} \hat{x}_{n,j} = 1, \hat{y}_n = 0, \hat{z}_n = 0, & \text{if } c_{n,j} = C_{n,\min}, \forall n \in \mathcal{N}, \\ \hat{x}_{n,j} = 0, \hat{y}_n = 1, \hat{z}_n = 0, & \text{if } c_{n,N+1} = C_{n,\min}, \forall n \in \mathcal{N}, \\ \hat{x}_{n,j} = 0, \hat{y}_n = 0, \hat{z}_n = 1, & \text{if } c_{n,N+2} = C_{n,\min}, \forall n \in \mathcal{N}, \end{cases} \quad (47)$$

where $c_{n,j} = (\rho/2) - \rho x_{n,j}^{[t+1]} - \lambda_{n,j}^{[t]}$, $c_{n,N+1} = (\rho/2) - \rho y_n^{[t+1]} - \mu_n^{[t]}$, and $c_{n,N+2} = (\rho/2) - \rho z_n^{[t+1]} - \gamma_n^{[t]}$ and $C_{n,\min}$ is the minimum

1. Initialization:
 $t = 1$, $\varepsilon = 0.01$, $\rho = 0.1$, $\lambda^{[t]} = 0$, and $t_{\max} = 2000$
 Choose an initial feasible solution $\hat{\Theta}^{[t]}$ satisfying (23)

2. Iteration:
while $\|\Theta - \hat{\Theta}\|_2 \geq \varepsilon$ and $t \leq t_{\max}$ **do**
 I) Global variable update: M computation units solve $P1'$
 and $P2'$ in parallel and a computation unit solves $P3'$ to
 update the global variables $\{\Theta\}^{[t+1]}$
 II) Local variable update: Update the local variables $\hat{\Theta}^{[t+1]}$ based on (47)
 III) Dual variable update: Update the dual variables $\{\lambda, \mu, \gamma\}^{[t+1]}$ based on the equations (48)-(50)
end while

3. Output:
 Output the optimal solution $\{x_{n,j}, y_n, z_n\}^*$.

ALGORITHM 2: Parallel optimization algorithm via ADMM.

value of all possible solutions for MD n , which is expressed as $C_{n,\min} = \min \{c_{n,j \in \mathcal{N}}, c_{n,N+1}, c_{n,N+2}\}$.

5.4. Dual Variable Update. After obtaining the global and local variables, we perform the process of dual variable update. In the $[t + 1]$ -th iteration, the dual variables are updated as follows:

$$\lambda_{n,j}^{[t+1]} = \lambda_{n,j}^{[t]} + \rho \left(x_{n,j}^{[t+1]} - \hat{x}_{n,j}^{[t+1]} \right), \quad \forall n \in \mathcal{N}, j \in \mathcal{N}, \quad (48)$$

$$\mu_n^{[t+1]} = \mu_n^{[t]} + \rho \left(y_n^{[t+1]} - \hat{y}_n^{[t+1]} \right), \quad \forall n \in \mathcal{N}, \quad (49)$$

$$\gamma_n^{[t+1]} = \gamma_n^{[t]} + \rho \left(z_n^{[t+1]} - \hat{z}_n^{[t+1]} \right), \quad \forall n \in \mathcal{N}. \quad (50)$$

In summary, the efficient offloading scheme is obtained by sequential iteration for global variables, local variables, and dual variables. By leveraging the ADMM algorithm to the problem P3, we firstly minimize the augmented Lagrangian function (25) over the global variables and decompose the optimization problem into some smaller subproblems, which is executed in parallel to improve the computing speed. Then, we performed the local variable update based on equation (47). Finally, the dual variables are updated based on equations (48)–(50). The algorithm is summarized in Algorithm 2.

6. Evaluation

In this section, we will present simulation results to verify the performance of the proposed D2D-enabled three-tier MEC networks. In our simulations, the number of BSs is 4. Each BS has a radius of 400 m. And there are 80 MDs uniformly distributed in the coverage of each BS. We assume that each MD can establish one wireless link with the BS and one D2D link with a neighboring MD. The maximum range of each D2D link is set as 50 m. The channel gain model of all wireless links is chosen as $h = 127 + 30 \log_{10} d$ (d in km), as suggested in [55]. Each MD has a task to be computed. The

TABLE 2: Simulation parameters.

Parameters	Value
The bandwidth resource B_n	20 MHz [20]
The number of BSs M	4
The transmission power of MDs P_n	20 dBm [4]
Power of background noise σ^2	-174 dBm/Hz [57]
Path-loss model	$127 + 30 * \log_{10} d$, d (km)
Local computational capability f_n^L	0.5GHz [56]
Local computation power P_n^L	24 dBm [20]
The computing capability of BSs F	20 GHz [38]
The computing capability of cloud F	100 GHz [33]

input data size of task n is $L_n \in [0.1, 4.0]$ Mbits. And the required computation amount of each task is assumed to be $C_n \in [2 \times 10^9, 5 \times 10^9]$ CPU cycles. The maximum delay tolerance of each task is 1 s. The computing capability of each MD is uniformly distributed from 0.5×10^9 to 2×10^9 CPU cycles/s. The maximum transmission power is set as 0.5 W [56] for each MD.

All random variables are independent for different MDs, modeling heterogeneous mobile computing resources. The main system parameters used in the simulation are summarized in Table 2, if not specified.

6.1. Performance Comparison. We first compare the performance of the proposed D2D-enabled three-tier MEC framework with the other two networks [23, 35, 58]:

- (1) In the D2D networks, tasks which can be cooperatively processed only among local MDs
- (2) In the MEC networks, vertical cooperation among local MD itself, edge node, and remote cloud node.
- (3) In the D2D-enabled MEC networks, where each task can be processed locally by MD itself or its neighboring MDs, by edge cloud, or by remote cloud.

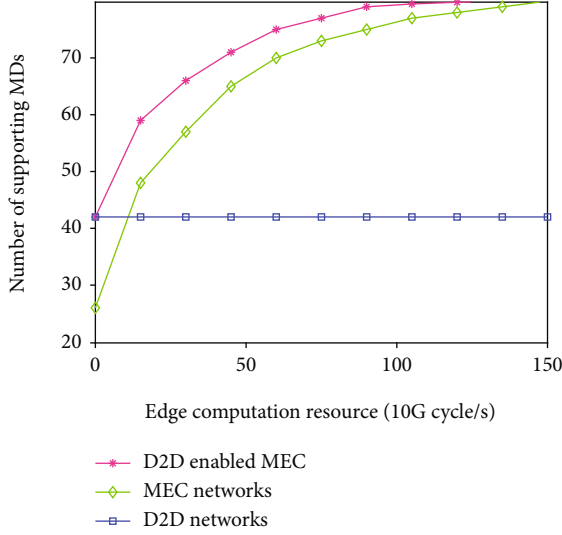


FIGURE 3: The number of supported MDs.

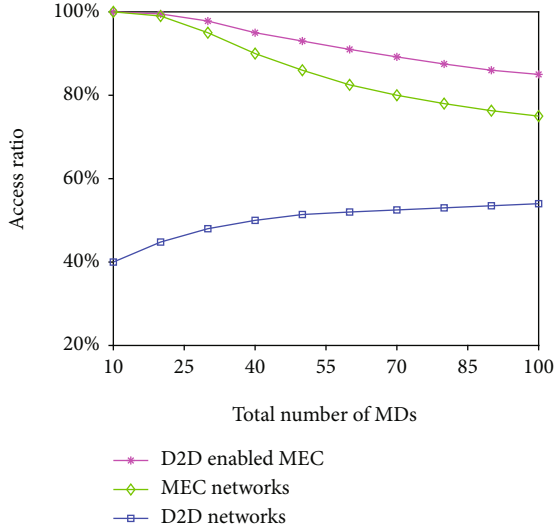


FIGURE 4: Access rate with different numbers of MDs.

From Figure 3, we can observe that the number of supporting MDs in these different networks to edge computing resources. First of all, the numbers of supporting MDs in the MEC networks and the D2D-enabled MEC networks both increases with the edge computing resources. While the number of supporting MDs in the D2D networks keeps invariable. The reason is that the D2D networks does not utilize the edge computing resources. Secondly, the D2D-enabled three-tier MEC networks achieves the best performance among these three systems. The reason is that the D2D-enabled MEC networks can utilize the computing resources of MDs, the edge servers, even the central cloud servers when the edge computing resource is insufficient. Moreover, the computing capacity of the D2D-enabled MEC networks is larger than that of the MEC networks since the computing resource of the MDs can be used for processing the offloaded computing tasks. Finally, the upper bounds

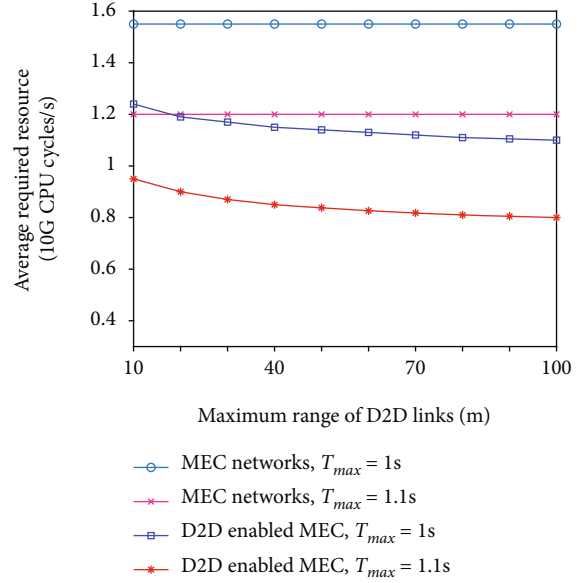


FIGURE 5: The average required edge computing resource vs. the maximum range of D2D links.

of the computing capacity of the D2D-enabled MEC networks and the MEC networks are the same, that is the overall number of MDs, i.e., 80. The result indicates that employing D2D communications can notably save the required computing resources from edge data center or cloud data center.

Figure 4 depicts the access ratio. It is the proportion of MDs that can be supported versus the total number of MDs. It can be observed that the access rates of both the MEC networks and D2D-enabled MEC networks decreasing with the number of MDs due to the limited edge computing resources, while the access rate of the D2D networks will correspondingly increase. The reason is that, the allocated edge computing resources for each MD in the MEC networks and D2D-enabled MEC networks decreases with the total number of MDs, which results in the smaller proportion of MDs that can be supported. Furthermore, the MDs computing resources can be fully utilized as the total number of MDs in the D2D networks increase, which eventually improves the access rate. We can also observe that the access rate of the D2D-enabled MEC networks decreases lower than that of the MEC networks since it utilizes the MD computing resources for capacity enhancement.

Figure 5 illustrates the average required computing resources for each MD, versus the maximum range of D2D links in the MEC networks and D2D-enabled three-tier MEC networks, respectively. It shows that the average required computing resources in the D2D-enabled MEC networks are smaller than those in the MEC networks, especially when the maximum range of D2D links becomes large. This is because each MD has more potential neighboring MDs to offload tasks as the range of D2D link becomes larger. In this way, the computing resources of all MDs can be further utilized via D2D offloading. We can also observe a reduced required computing resource with the decreased delay tolerance in both networks.

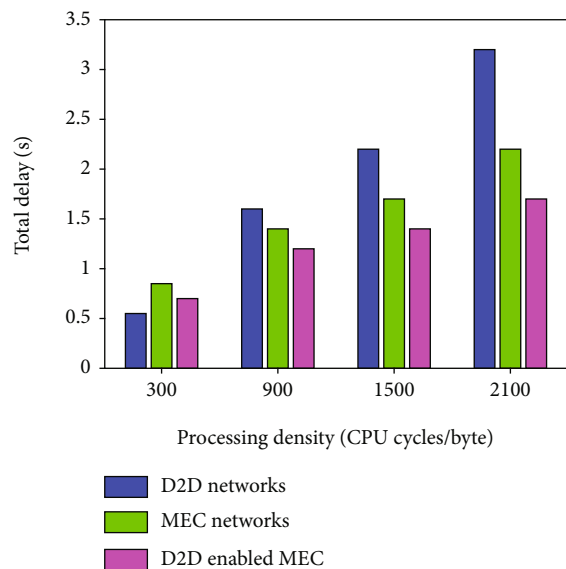


FIGURE 6: Total delay with different processing densities.

Figure 6 presents the total delay with different computation processing densities. Comparing the proposed scheme with two benchmarks (i.e., MEC networks and D2D networks), we can observe the following: when the processing density is low, the task processed at the device tier performs well compared to offloading to the edge cloud or central cloud DC. This is because the transmission latency on account of offloading from the MD to the edge node or cloud node dominates over the local processing time at the local device tier.

Nevertheless, as the processing density increases, the proposed scheme outperforms the D2D processing at the device tier. Furthermore, with higher processing density, only the MEC networks may not be able to meet the low-latency requirement. On the other hand, as our proposed D2D-enabled MEC networks optimize the offloading decisions, we obtain a significantly lower total latency compared to the other solutions.

7. Conclusion

In this article, we consider cooperative offloading in D2D-enabled three-tier MEC networks as a part of the IoT. By leveraging the advantages of both MEC and D2D cooperation techniques, the computing resources in edge nodes, cloud nodes, and D2D MDs can be effectively utilized. We jointly optimize the task offloading decisions and computing resource allocation to minimize the total delay of tasks. The formulated problem is a mixed-integer nonlinear optimization problem. When networks grow larger, it is difficult to be solved. To work out this problem, we adopt a parallel optimization framework based on the ADMM algorithm. Simulation results show that the proposed framework and solution outperform the other solutions with various parameters. In our future work, we may consider the queue of tasks, partible offloading, and privacy protection when investigating the offloading problem.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the National Nature Science Foundation of China Projects (No. 61871051).

References

- [1] M. Jia, J. Cao, and W. Liang, "Optimal cloudlet placement and user to cloudlet allocation in wireless metropolitan area networks," *IEEE Transactions on Cloud Computing*, vol. 5, no. 4, pp. 725–737, 2017.
- [2] T. X. Tran, A. Hajisami, P. Pandey, and D. Pompili, "Collaborative mobile edge computing in 5G networks: new paradigms, scenarios, and challenges," *IEEE Communications Magazine*, vol. 55, no. 4, pp. 54–61, 2017.
- [3] L. Liu, Z. Chang, X. Guo, S. Mao, and T. Ristaniemi, "Multi-objective optimization for computation offloading in fog computing," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 283–294, 2017.
- [4] X. Chen, L. Jiao, W. Li, and X. Fu, "Efficient multi-user computation offloading for mobile-edge cloud computing," *IEEE/ACM Transactions on Networking*, vol. 24, no. 5, pp. 2795–2808, 2016.
- [5] N. Chen, B. Rong, X. Zhang, and M. Kadoch, "Scalable and flexible massive MIMO precoding for 5G H-CRAN," *IEEE Wireless Communications*, vol. 24, no. 1, pp. 46–52, 2017.
- [6] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: the communication perspective," *IEEE Communication Surveys and Tutorials*, vol. 19, no. 4, pp. 2322–2358, 2017.
- [7] K. Zhang, Y. Mao, S. Leng et al., "Energy-efficient offloading for mobile edge computing in 5G heterogeneous networks," *IEEE Access*, vol. 4, pp. 5896–5907, 2016.
- [8] M. Li, Q. Wu, J. Zhu, R. Zheng, and M. Zhang, "A computing offloading game for mobile devices and edge cloud servers," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 2179316, 10 pages, 2018.
- [9] E. Cuervo, A. Balasubramanian, D.-k. Cho et al., "Maui: making smartphones last longer with code offload," in *Proceedings of the 8th international conference on Mobile systems, applications, and services - MobiSys '10*, pp. 49–62, San Francisco, California, USA, June, 2010.
- [10] Ning Zhang, Nan Cheng, A. T. Gamage, Kuan Zhang, J. W. Mark, and Xuemin Shen, "Cloud assisted hetnets toward 5G wireless networks," *IEEE Communications Magazine*, vol. 53, 6-Supplement, pp. 59–65, 2015.
- [11] M. V. Barbera, S. Kosta, A. Mei, and J. Stefa, "To offload or not to offload? the bandwidth and energy costs of mobile cloud computing," in *2013 Proceedings IEEE INFOCOM*, Turin, Italy, 2013.
- [12] R. Deng, R. Lu, C. Lai, T. H. Luan, and H. Liang, "Optimal workload allocation in fog-cloud computing towards balanced

- delay and power consumption,” *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1171–1181, 2016.
- [13] P. Mach and Z. Becvar, “Mobile edge computing: a survey on architecture and computation offloading,” *IEEE Communication Surveys and Tutorials*, vol. 19, no. 3, pp. 1628–1656, 2017.
- [14] A. Mukherjee, D. de, and D. G. Roy, “A power and latency aware cloudlet selection strategy for multi-cloudlet environment,” *IEEE Transactions on Cloud Computing*, vol. 7, no. 1, pp. 141–154, 2019.
- [15] S. Sardellitti, G. Scutari, and S. Barbarossa, “Joint optimization of radio and computational resources for multicell mobile-edge computing,” *IEEE Transactions on Signal and Information Processing over Networks*, vol. 1, no. 2, pp. 89–103, 2015.
- [16] A. Ewaisha and C. Tepedelenioglu, “Offloading deadline-constrained cellular traffic,” in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, pp. 1447–1451, Pacific Grove, CA, USA, October 2018.
- [17] Y. Xiao and M. Krunz, “QoE and power efficiency tradeoff for fog computing networks with fog node cooperation,” in *INFOCOM*, pp. 1–9, Atlanta, GA, USA, May 2017.
- [18] Y. Yu, J. Zhang, and K. B. Letaief, “Joint subcarrier and CPU time allocation for mobile edge computing,” in *IEEE GLOBECOM*, pp. 1–6, Washington, DC, USA, December 2016.
- [19] Y. Mao, J. Zhang, and K. B. Letaief, “Dynamic computation offloading for mobile-edge computing with energy harvesting devices,” *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3590–3605, 2016.
- [20] C. Wang, C. Liang, F. R. Yu, Q. Chen, and L. Tang, “Computation offloading and resource allocation in wireless cellular networks with mobile edge computing,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 8, pp. 4924–4938, 2017.
- [21] T. X. Tran and D. Pompili, “Joint task offloading and resource allocation for multi-server mobile-edge computing networks,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 856–868, 2019.
- [22] K. Kumar and Yung-Hsiang Lu, “Cloud computing for mobile users: can offloading computation save energy?,” *Computer*, vol. 43, no. 4, pp. 51–56, 2010.
- [23] Y. He, J. Ren, G. Yu, and Y. Cai, “Joint computation offloading and resource allocation in D2D enabled MEC networks,” in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, Shanghai, China, May 2019.
- [24] J. Feng, L. Zhao, D. Jianbo, X. Chu, and F. R. Yu, “Computation offloading and resource allocation in D2D-enabled mobile edge computing,” in *2018 IEEE International Conference on Communications (ICC)*, Kansas City, MO, USA, May 2018.
- [25] X. Chen, L. Pu, L. Gao, W. Wu, and D. Wu, “Exploiting massive D2D collaboration for energy-efficient mobile edge computing,” *IEEE Wireless Communications*, vol. 24, no. 4, pp. 64–71, 2017.
- [26] S. Josilo and G. Dan, “Decentralized algorithm for randomized task allocation in fog computing systems,” *IEEE/ACM Transactions on Networking*, vol. 27, no. 1, pp. 85–97, 2019.
- [27] C. Zheng, Z. Zhou, T. Ristaniemi, and Z. Niu, “Energy efficient optimization for computation offloading in fog computing system,” in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, Singapore, December 2017.
- [28] Y. Hao, M. Chen, L. Hu, M. S. Hossain, and A. Ghoneim, “Energy efficient task caching and offloading for mobile edge computing,” *IEEE Access*, vol. 6, pp. 11365–11373, 2018.
- [29] H. Huang, J. Mu, and X. Jing, “Cooperative spectrum sensing based on centralized double threshold in mcn,” *China Communications*, vol. 17, no. 5, pp. 235–242, 2020.
- [30] C. Wang, C. Liang, F. R. Yu, Q. Chen, and L. Tang, “Joint computation offloading, resource allocation and content caching in cellular networks with mobile edge computing,” in *2017 IEEE International Conference on Communications (ICC)*, Paris, France, May 2017.
- [31] J. Mu, X. Jing, J. Xie, and Y. Zhang, “Multistage spectrum sensing scheme with SNR estimation,” *IET Communications*, vol. 13, no. 9, pp. 1148–1154, 2019.
- [32] B. Ji, Z. Chen, S. Chen, B. Zhou, C. Li, and H. Wen, “Joint optimization for ambient backscatter communication system with energy harvesting for IoT,” *Mechanical Systems and Signal Processing*, vol. 135, pp. 106412–106412.10, 2020.
- [33] T. Q. Thanh, J. Tang, Q. D. la, and T. Q. S. Quek, “Offloading in mobile edge computing: task allocation and computational frequency scaling,” *IEEE Transactions on Communications*, vol. 65, no. 8, pp. 3571–3584, 2017.
- [34] B. Rong, Y. Qian, and K. Lu, “Integrated downlink resource management for multiservice wimax networks,” *IEEE Transactions on Mobile Computing*, vol. 6, no. 6, pp. 621–632, 2007.
- [35] C. You, K. Huang, H. Chae, and B.-H. Kim, “Energy-efficient resource allocation for mobile-edge computation offloading,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1397–1411, 2017.
- [36] Z. Liu, J. Zhang, and W. Jingyan, “Joint optimization of server placement and content caching in mobile edge computing networks,” in *Proceedings of the 2019 8th International Conference on Networks, Communication and Computing*, pp. 149–153, Luoyang, China, December 2019.
- [37] M. Mukherjee, S. Kumar, M. Shojafar, Q. Zhang, and C. X. Mavromoustakis, “Joint task offloading and resource allocation for delay-sensitive fog networks,” in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, pp. 1–7, Shanghai, China, May 2019.
- [38] Y. Wang, X. Tao, X. Zhang, P. Zhang, and Y. T. Hou, “Cooperative task offloading in three-tier mobile computing networks: an admm framework,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 2763–2776, 2019.
- [39] T. Liang, L. Yong, and W. Gao, “A hierarchical edge cloud architecture for mobile computing,” in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, pp. 1–9, San Francisco, CA, USA, April 2016.
- [40] J. du, L. Zhao, J. Feng, and X. Chu, “Computation offloading and resource allocation in mixed fog/cloud computing systems with min-max fairness guarantee,” *IEEE Transactions on Communications*, vol. 66, no. 4, pp. 1594–1608, 2018.
- [41] H. Xing, L. Liu, J. Xu, and A. Nallanathan, “Joint task assignment and resource allocation for D2D-enabled mobile-edge computing,” *IEEE Transactions on Communications*, vol. 67, no. 6, pp. 4193–4207, 2019.
- [42] C. Kachris and I. Tomkos, “A survey on optical interconnects for data centers,” *IEEE Communications Surveys & Tutorials*, vol. 14, no. 4, pp. 1021–1036, 2012.
- [43] Z. Liu, J. Zhang, Y. Li, L. Bai, and Y. Ji, “Joint jobs scheduling and lightpath provisioning in fog computing micro datacenter networks,” *Journal of Optical Communications and Networking*, vol. 10, no. 7, pp. B152–B163, 2018.

- [44] Y. Ji, R. Gu, Z. Yang, J. Li, H. Li, and M. Zhang, "Artificial intelligence-driven autonomous optical networks: 3S architecture and key technologies," *Science China Information Sciences*, vol. 63, no. 6, 2020.
- [45] I. Komnios, F. Tsapeli, and S. Gorinsky, "Cost-effective multi-mode offloading with peer-assisted communications," *Ad Hoc Networks*, vol. 25, pp. 370–382, 2015.
- [46] Y. Cui, J. Song, K. Ren et al., "Software defined cooperative offloading for mobile cloudlets," *IEEE/ACM Transactions on Networking*, vol. 25, no. 3, pp. 1746–1760, 2017.
- [47] X. Chen, "Decentralized computation offloading game for mobile cloud computing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 4, pp. 974–983, 2015.
- [48] X. Lyu, H. Tian, C. Sengul, and P. Zhang, "Multiuser joint task offloading and resource optimization in proximate clouds," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3435–3447, 2017.
- [49] C. Sonmez, A. Ozgovde, and C. Ersoy, "Edgecloudsim: an environment for performance evaluation of edge computing systems," *Transactions on Emerging Telecommunications Technologies*, vol. 29, no. 11, 2018.
- [50] S. P. Boyd, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.
- [51] H. K. Nguyen, Y. Zhang, Z. Chang, and Z. Han, "Parallel and distributed resource allocation with minimum traffic disruption for network virtualization," *IEEE Transactions on Communications*, vol. 65, no. 3, pp. 1162–1175, 2017.
- [52] E. Che, H. D. Tuan, and H. H. Nguyen, "Joint optimization of cooperative beamforming and relay assignment in multi-user wireless relay networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 10, pp. 5481–5495, 2014.
- [53] R. Horst and N. V. Thoai, "DC programming: overview," *Journal of Optimization Theory and Applications*, vol. 103, no. 1, pp. 1–43, 1999.
- [54] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge Univ. Press, Cambridge U. K., 2004.
- [55] M. Chen and Y. Hao, "Task offloading for mobile edge computing in software defined ultra-dense network," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 3, pp. 587–597, 2018.
- [56] M. H. Chen, B. Liang, and M. Dong, "Multi-user multi-task offloading and resource allocation in mobile cloud systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 10, pp. 6790–6805, 2018.
- [57] Y. Mao, J. Zhang, and K. B. Letaief, "Power-delay tradeoff in multi-user mobile-edge computing systems," in *2016 IEEE Global Communications Conference (GLOBECOM)*, Washington, DC, USA, December 2016.
- [58] Y. He, J. Ren, G. Yu, and Y. Cai, "D2D communications meet mobile edge computing for enhanced computation capacity in cellular networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 3, pp. 1750–1763, 2019.

Research Article

CNN and DCGAN for Spectrum Sensors over Rayleigh Fading Channel

Junsheng Mu ¹, Youheng Tan,² Dongliang Xie,¹ Fangpei Zhang,³ and Xiaojun Jing ²

¹The State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China

²School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China

³Information Science Research Institute of China Electronics Technology Group Corporation, Beijing 100876, China

Correspondence should be addressed to Xiaojun Jing; jxiaojun@bupt.edu.cn

Received 16 March 2021; Accepted 15 July 2021; Published 10 August 2021

Academic Editor: Ding Xu

Copyright © 2021 Junsheng Mu 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.

Spectrum sensing (SS) has attracted much attention in the field of Internet of things (IoT) due to its capacity of discovering the available spectrum holes and improving the spectrum efficiency. However, the limited sensing time leads to insufficient sampling data due to the tradeoff between sensing time and communication time. In this paper, deep learning (DL) is applied to SS to achieve a better balance between sensing performance and sensing complexity. More specifically, the two-dimensional dataset of the received signal is established under the various signal-to-noise ratio (SNR) conditions firstly. Then, an improved deep convolutional generative adversarial network (DCGAN) is proposed to expand the training set so as to address the issue of data shortage. Moreover, the LeNet, AlexNet, VGG-16, and the proposed CNN-1 network are trained on the expanded dataset. Finally, the false alarm probability and detection probability are obtained under the various SNR scenarios to validate the effectiveness of the proposed schemes. Simulation results state that the sensing accuracy of the proposed scheme is greatly improved.

1. Introduction

In recent years, the spectrum resource has been more and more scarce due to the great demand for wireless communication, Internet of Things (IoT), Artificial Intelligence (AI) [1–3], etc. One of the most important issues of wireless communication technology is to improve its spectrum efficiency in the near future. As a possible scheme to improve spectrum efficiency, cognitive radio (CR) [4] has attracted much attention.

The core idea behind CR is to realize dynamic spectrum allocation (DSA) and spectrum sharing by spectrum sensing and the intelligent learning ability of the system [5]. The most important technology behind CR is to periodically monitor the absence or the presence of the registered users within the observed bands, named spectrum sensing (SS) [6]. In SS, the registered users are the primary users (PU) of the observed bands and have the priority to the spectrum. The purpose of CR is to opportunistically access the registered spectrum when PU is absent. As a result, the cognitive users are the secondary

users (SU). Once the PU is back, the SU will release the spectrum at once and wait for the other opportunity.

Classical SS contains matched filtering, energy detector (ED) [7], cyclic spectrum detection [8], covariance matrix detection [9], etc. The sensing performance of the matched filtering is optimal if the prior knowledge of the primary signal is known in advance. ED is the optimal blind detector considering both sensing performance and sensing complexity. However, it suffers from noise uncertainty under the low signal-to-noise ratio (SNR) regimes. The sensing performance of cyclic spectrum detection and covariance matrix detection is improved in the low-SNR case compared with ED at the expense of a higher complexity. However, these traditional SS schemes either have poor performance or have high complexity.

Recently, with the wide application of the deep learning (DL) in the field of computer vision [10], the wireless communication based on DL has been a hot topic [11–13]. The essence of DL is to provide a method of automatically learning

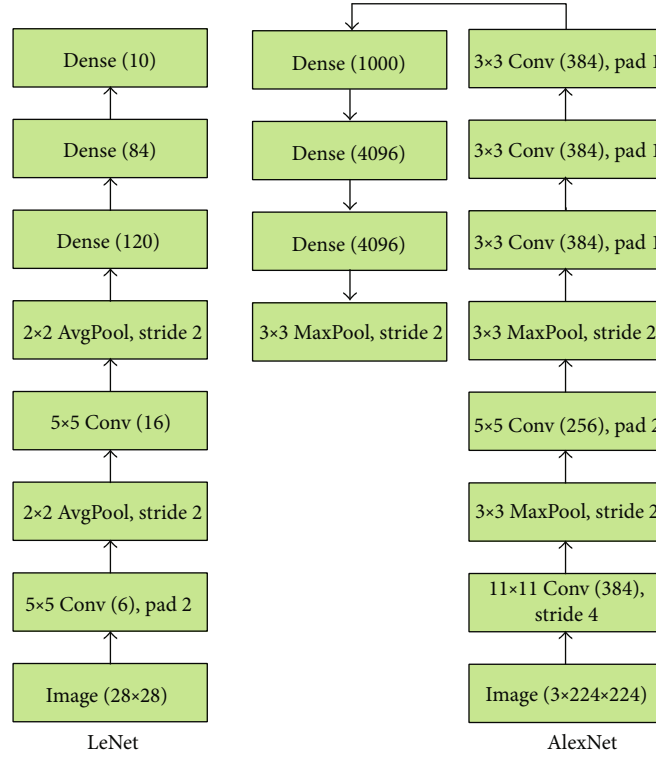


FIGURE 1: The network architecture of LeNet and AlexNet.

pattern features and combine the features, thus, reducing the incompleteness caused by artificial design features. In [14], a stacked autoencoder based spectrum sensing approach (SAE-SS) was proposed to relieve the influence from carrier frequency offset (CFO), timing delay, and noise uncertainty. A deep learning based signal detector was considered to exploit the underlying structural information of the modulated signals in [15]. The transfer learning strategies were used in [16] to improve the performance for real-world signals. In [17], the convolutional neural network- (CNN-) long short term memory network (LSTM) detector was proposed to extract the spatial and temporal features of the input.

Motivated by the mentioned above, DL is applied to SS in this paper, where the covariance matrix of the received signal is converted into the true color picture. Then, an improved deep convolutional generative adversarial network (DCGAN) is proposed to expand the training set for the issue of data shortage.

After that, the LeNet [18], AlexNet [19], VGG-16 [20], and a novel network are trained based on the extended data. Finally, the simulations are made to validate the effectiveness of the proposed schemes. The main contributions of this paper are concluded as follows.

- (1) The two-dimensional dataset of the received signal is established under the various SNR conditions. Each SNR contains 4000 samples from -10 dB to 2 dB
- (2) An improved DCGAN network is proposed to expand the obtained two-dimensional dataset. In

the expanded dataset, each SNR contains 8000 samples

- (3) The LeNet, AlexNet, and VGG-16 networks are trained on the expanded dataset. The corresponding false alarm probability and detection probability are given under the various SNR scenarios
- (4) Based on the sensing performance of the LeNet, AlexNet, and VGG-16 networks, an improved network is provided in this paper to balance the sensing performance and the sensing complexity

The remainder of the paper is organized as follows. Section 2 introduces the related work and gives the system model. In Section 3, the improved DCGAN scheme is discussed to solve the issue of data shortage. The SS with the LeNet, AlexNet, VGG-16, and an improved network is conducted. Finally, conclusions are drawn in Section 5.

2. Related Work

In this section, three classical convolutional neural networks (CNN) and the deep convolutional generative adversarial network (DCGAN) [21] are reviewed. In addition, the system model of this paper is provided.

2.1. CNN. CNN is a feed-forward neural network. This network model uses a gradient descent method to minimize the loss function and reversely adjusts the weight parameters

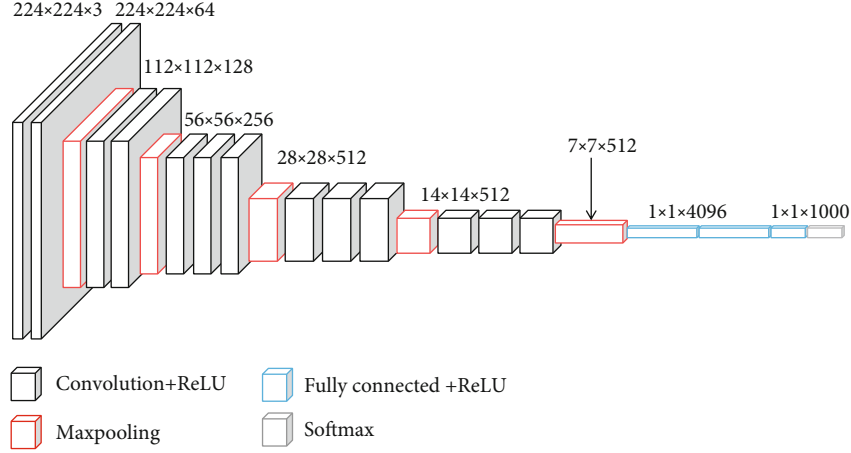


FIGURE 2: The network architecture of VGG-16.

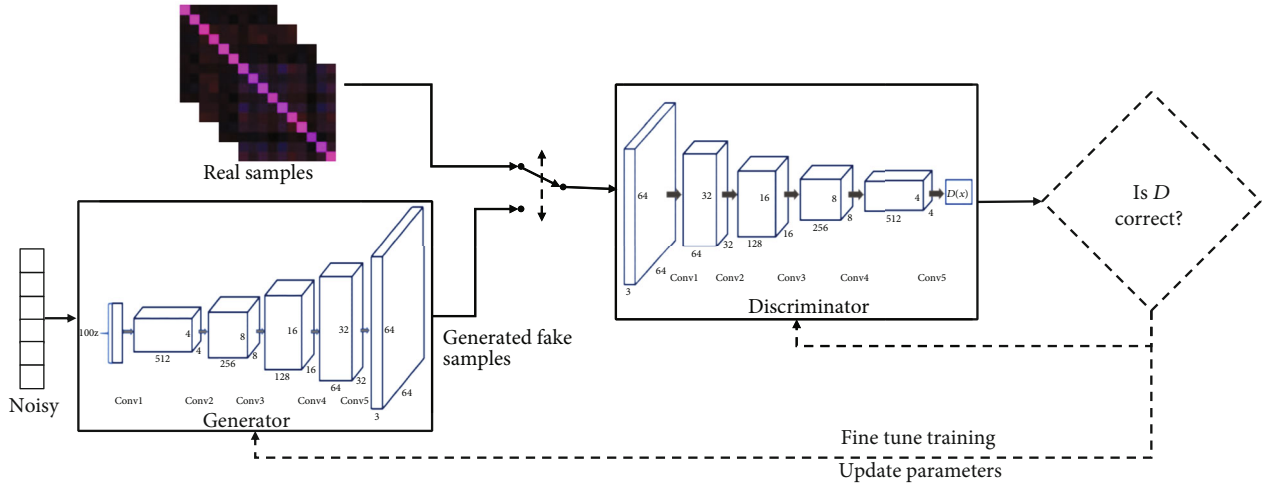


FIGURE 3: Basic framework of DCGAN.

in the network layer by layer. Three classical CNN networks are considered in this paper, LeNet, AlexNet, and VGG-16.

As it is shown in Figure 1, the LeNet network contains three convolutional layers with the size 5×5 , two pooling layers with the size 2×2 , and three fully connected layers [18]. The AlexNet network contains five convolutional layers, three pooling layers with the size 3×3 , and three fully connected layers [19].

As it is shown in Figure 2, the VGG-16 network contains 13 convolutional layers with the size 3×3 and three fully connected layers [20]. The maximum pooling is considered in VGG-16 network with the size 2×2 . The first two pooling layers are followed by two convolutional layers while the rest pooling layers are followed by three convolutional layers.

2.2. DCGAN. The generative adversarial networks (GAN) [21] have attracted wide attention in the field of machine learning because of its great potential to imitate high-dimensional and complex real data. For scenarios where there is a lack of data, it can be used to generate more sample data. In order to solve the problem of high acquisition cost of training set samples, this paper utilizes GAN to generate

more training set samples. Generative adversarial networks include generating network (Generator) and discriminating network (Discriminator). The generating network learns the real data distribution to generate new data under the guidance of the discriminating network. Deep convolutional generative adversarial networks are one of the more effective and stable networks based on GAN. The basic framework of DCGAN [22] is shown in Figure 3.

Assume the input of the Generator is the random Gaussian noise z and its output is the fake sample $G(z)$. The true sample x and fake sample $G(z)$ are input to the Discriminator, respectively, and the corresponding outputs are $D(x)$ and $D(G(z))$. $D(x)$ denotes the probability that the input x of the Discriminator is a real sample. The Discriminator is to make $D(x)$ tend to 1 and $D(G(z))$ tend to 0. At the same time, the Generator is to make $D(G(z))$ close to 1. The loss function of DCGAN is shown as

$$L(D, G) = E_{x \sim P_r} [\log D(x)] + E_{z \sim P_g} [\log (1 - D(x))], \quad (1)$$

where P_r represents the real sample distribution and P_g represents the fake sample distribution.

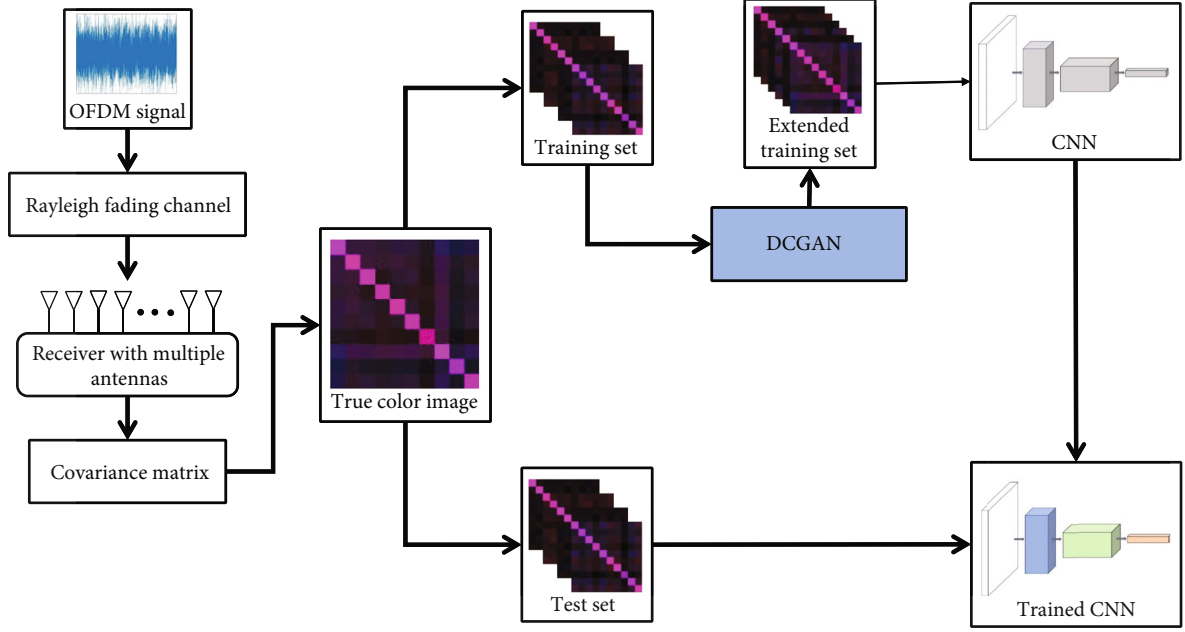


FIGURE 4: System model.

The objective functions of the Discriminator and the Generator are, respectively, written as

$$\max_D L(D, G), \quad (2)$$

$$\max_G \max_D L(D, G). \quad (3)$$

(3) can be further rewritten as

$$\max_G \max_D E_{x \sim p_g} [\log(1 - D(x))]. \quad (4)$$

2.3. System Model. In this paper, the signal is received by a multiantenna system, and then the covariance matrix of the signal is calculated. After that, the covariance matrix is transformed into a true color graph as a data set. The system model of this paper is shown in Figure 4.

An M -element antenna system is considered to receive the signals based on N observation vectors. Let $s_i(n)$, $n = 0, 1, \dots, N-1$ denote the n -th discrete-time sample at the i -th antenna. Generally, the spectrum sensing can be regarded as a binary classification [21],

$$s_i(n) = \begin{cases} w_i(n), & H_0, \\ x_i(n) + w_i(n), & H_1, \end{cases} \quad (5)$$

where $w_i(n)$ denotes the background noise and $x_i(n)$ denotes primary signal vector with the Rayleigh fading [7]. H_0 and H_1 , respectively, signify the absence and the presence of PU.

Let $s_i = [s_i(0), s_i(1), \dots, s_i(N-1)]$ denote the sampling sequence at the i -th antenna and its average can be expressed as

$$s_{i-a} = \frac{1}{N} \sum_{n=0}^{N-1} s_i(n). \quad (6)$$

The time series matrix of the received signal S can be formulated as

$$S = \begin{bmatrix} s_1(0) - s_{1-a} & s_1(1) - s_{1-a} & \cdots & s_1(N) - s_{1-a} \\ s_2(0) - s_{2-a} & s_2(1) - s_{2-a} & \cdots & s_2(N) - s_{2-a} \\ \vdots & \vdots & \cdots & \vdots \\ s_M(0) - s_{M-a} & s_M(1) - s_{M-a} & \cdots & s_M(N) - s_{M-a} \end{bmatrix}. \quad (7)$$

The corresponding covariance matrix is

$$R(N) = \frac{1}{N} S \times S^H. \quad (8)$$

For the real part and the imaginary part of S , the covariance matrix is considered as the two channels of the true color image, together with a zero matrix as the third channel of the true color image.

The samples are divided into training set and test set. Then, the training set is expanded by DCGAN. After that, CNN is trained by the expanded training set. Finally, the sensing performance is validated by the test set.

The false alarm probability and the detection probability can be formulated as

$$P_d = P\{\varphi[R(N)] > \lambda | H_1\}, \quad (9)$$

$$P_f = P\{\varphi[R(N)] > \lambda | H_0\}, \quad (10)$$

where $\varphi[R(N)]$ denotes the feature extraction operation of the proposed network and λ denotes the sensing threshold.

3. Data Enhancement with DCGAN

In this section, the data enhancement scheme with DCGAN is discussed and an improved DCGAN scheme is proposed, where the python3.7 and pytorch1.5 machine learning libraries are used to implement generative adversarial network and convolutional neural network. The hardware CPU is Inter(R) Core(TM) i5-6300HQ and the GPU is NVIDIA GeForce GTX 960 M.

3.1. Data Generation. In this subsection, the original training set and test set are generated, where the OFDM signal [23] is considered as the primary signal, and the Rayleigh fading [24] is regarded as the propagation channel. The size of the sampling covariance matrix is $10 \times 10 \times 2$. The number of antennas in the multiantenna system used in this paper is $M = 10$. The channel number of a true color image is 3, where the channel number of the sampling covariance matrix is 2 (the real part and the imaginary part), and the third channel is set to a zero matrix. The matrix size depends on the antenna number of the multiantenna system.

In the process of data generation, 8 datasets are generated, whose SNR varies from -10 dB to 2 dB with the step 2 dB. Each dataset is divided into two parts, H_0 and H_1 . For the H_0 part, the real and imaginary parts are, respectively, sampled to obtain 3000 sets of data, where each set of data sampling points $N = 1000$ and contains two matrices. Then, this set of data is a $10 \times 1000 \times 2$ sampling time series matrix. Calculate the sampling covariance matrix of each matrix according to (7). Then, 3000 sets of dual-channel sampling covariance matrices can be obtained. After that, these 3000 sets of data are randomly sorted and converted into true color images, where the first channel of the true color image is the first channel of the sampling covariance matrix, the second channel is the zero matrix, and the third channel is the second channel of the sampling covariance matrix. Finally, take the first 1000 groups as the data of the H_0 category in the test set, and take the last 2000 groups as the data of the H_0 category in the original training set.

For the H_1 part, the same operations are conducted as the H_0 . As a result, the 1000 sets of test set data and 2000 sets of original training set data for the H_1 part.

Figure 5 exhibits the obtained true color image in the H_0 and H_1 case under 2 dB, where the left image corresponds with the H_0 case and the right image corresponds with the H_1 case. From Figure 5, the color of the H_0 case is very dark except for the diagonal, which indicates that the value of the corresponding two-channel sampling covariance matrix is very small. While in the H_1 case, the color of the image is

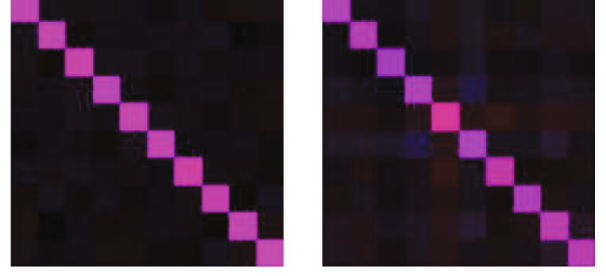


FIGURE 5: True color image in the H_0 and H_1 cases under 2 dB.

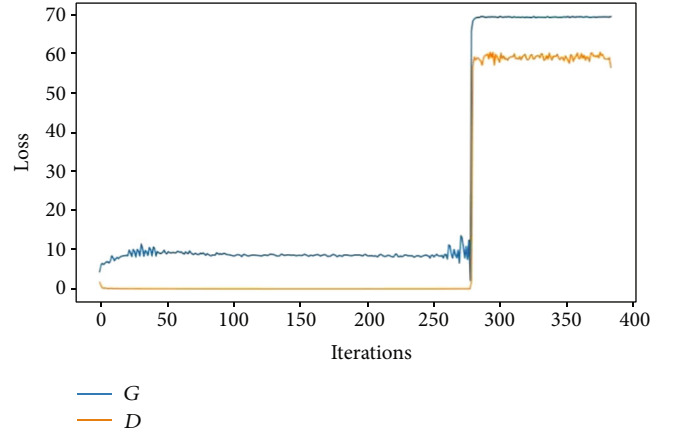


FIGURE 6: The loss of the training.

uneven and somewhat chaotic, which indicates that the corresponding two-channel sampling covariance matrix values are also chaotic.

3.2. Data Enhancement. The sampled data is enhanced in this subsection based on the classical DCGAN scheme, where the sampled data is doubled, from 4000 images to 8000 images for each SNR.

In Figure 6, the loss of the training with the iterations is exhibited, where G denotes the Generator, D denotes the Discriminator, and the number of training cycles is 16. According to Figure 6, when it runs to the 10th cycle, a large loss occurs due to the gradient explosion. As a result, how to effectively reduce the loss determines the quality of data enhancement.

In Figure 7, the initial weights of the convolution kernel are adjusted, from the Gaussian distribution with the mean 0 and the variance 0.02 to the Gaussian distribution with the mean 0.05 and the variance 0.02. As a result, the loss function is enlarged, and the trend of the loss function can be better observed. From Figure 7, to increase the mean of the initial weight will not only increase the loss function value but also slow down the training speed and reduce the training gradient.

Figures 8 and 9, respectively, exhibit the training result with the mean of initial weight 0.05 and 0.02, where the left image denotes the original image and the right image denotes the generated image with DCGAN. From Figures 8 and 9, the smaller the mean of initial weight is, the better the generated image quality is. It can be seen that the initial weight average still has a great influence on the convergence speed of the model. The larger the initial weight of the convolution kernel,

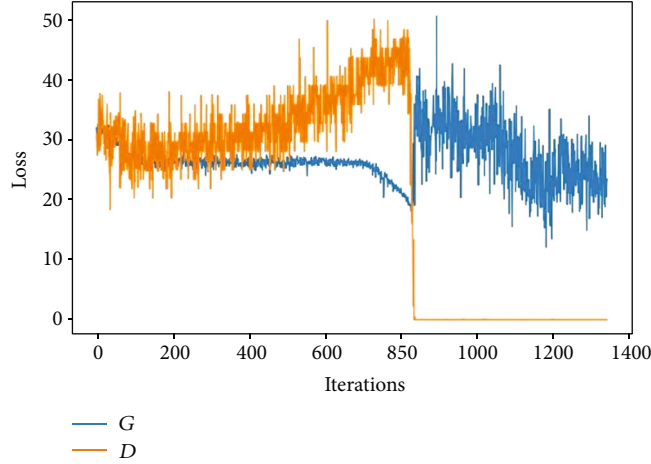


FIGURE 7: The training loss when the mean of initial weight is 0.05.

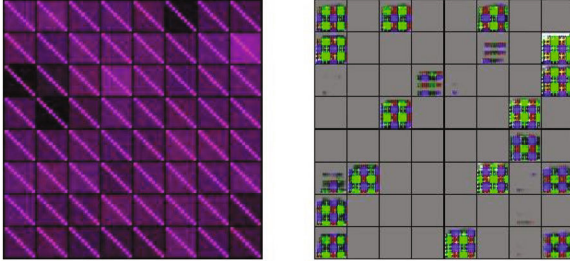


FIGURE 8: 82 times training result with the mean of initial weight 0.05.

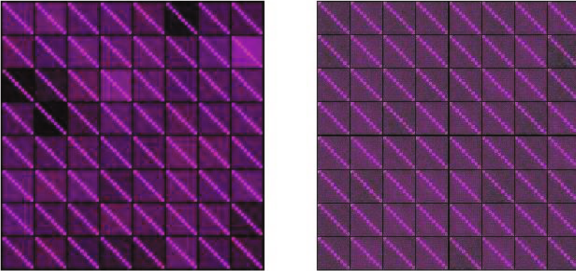


FIGURE 9: 82 times training result with the mean of initial weight 0.02.

the slower the convergence speed may be, and the slower the gradient will be. But at the same time, more useful information can be obtained.

3.3. Improved DCGAN. According to the discussion mentioned above, to increase the mean of the initial weight reduces the training gradient. However, it sacrifices the generated image quality. The improved DCGAN scheme is made to solve the above issue in this subsection.

The change of $D(G(z))$ before and after the gradient explosion is shown in Figure 10 with the mean of initial weight 0. From Figure 10, when the gradient explosion occurs, the loss functions of the generating network and the discriminating network become extremely large while $D(x)$

[0/16]	Loss_D: 16918	Loss_G: 4.3525	$D(x)$: 0.6500	$D(G(x))$: 0.6606/0.175
[2/16]	Loss_D: 0.0087	Loss_G: 9.4516	$D(x)$: 0.9994	$D(G(x))$: 0.0081/0.0001
[4/16]	Loss_D: 0.0025	Loss_G: 8.8306	$D(x)$: 0.9996	$D(G(x))$: 0.0020/0.0002
[6/16]	Loss_D: 0.0028	Loss_G: 8.6410	$D(x)$: 0.9996	$D(G(x))$: 0.0024/0.0002
[8/16]	Loss_D: 0.0034	Loss_G: 8.3489	$D(x)$: 0.9988	$D(G(x))$: 0.0022/0.0003
[10/16]	Loss_D: 58.5886	Loss_G: 69.2882	$D(x)$: 0.0000	$D(G(x))$: 0.0000/0.0000
[12/16]	Loss_D: 58.5617	Loss_G: 69.2911	$D(x)$: 0.0000	$D(G(x))$: 0.0000/0.0000

FIGURE 10: The operation status comparisons before and after the gradient explosion with the mean of initial weight 0.

and $D(G(z))$ are both 0. At this point, both networks will be collapsed. Motivated by this, an adjustment scheme is proposed as it is shown in Table 1.

Figures 11 and 12 show the improved DCGAN result with the mean of initial weight 0 and the improved DCGAN loss with the mean of initial weight 0 when the algorithm runs to the 40th training loop. It can be seen that the generated map has been able to reach the point of being fake. Moreover, from the point of view of the loss function, since the 100th training, the generating network and the discriminating network have also shown an obvious upward and downward jitter trend, that is, adversarial evolution. At the same time, it can be concluded that the mean value of the initialization parameter weight changes from 0.2 to 0, which speeds up the convergence of the model.

4. SS with CNN Network

For SS, two factors determine the sensing performance, detection probability (PD), and false alarm probability (PFA). As a result, the PD and PFA are provided with LeNet, AlexNet, VGG-16, and the proposed CNN-1 network in this section.

4.1. LeNet. Figures 13 and 14, respectively, show the PD [25] and PFA [26] under various SNR conditions. From Figures 13 and 14, the detection performance of the LeNet network is obviously different under different SNRs. When the SNR is small, PD is small and PFA is high due to the obtained nonsignificant features. However, when $\text{SNR} \geq -4$ dB, PD can be maintained at or above 0.9, and PFA is basically lower than 0.1. In addition, after data enhancement with the proposed scheme, the mean value of PD becomes higher,

TABLE 1: The adjustment scheme of DCGAN.

	$D(G(z))$	Emergency action
Emergency action of judgement network	$0.50 < D(G(z)) \leq 0.65$	2 extra training times
	$0.65 < D(G(z)) \leq 0.75$	3 extra training times
	$0.75 < D(G(z)) \leq 0.85$	4 extra training times
	$0.85 < D(G(z)) \leq 0.95$	5 extra training times
	$0.95 < D(G(z))$	6 extra training times
Emergency action of judgement network	$0.14 < D(G(z)) \leq 0.19$	1 extra training time
	$0.10 < D(G(z)) \leq 0.14$	2 extra training times
	$0.07 < D(G(z)) \leq 0.10$	3 extra training times
	$D(G(z)) \leq 0.07$	4 extra training times

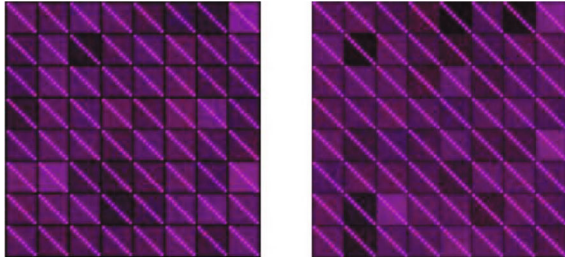


FIGURE 11: Improved DCGAN result with the mean of initial weight 0.

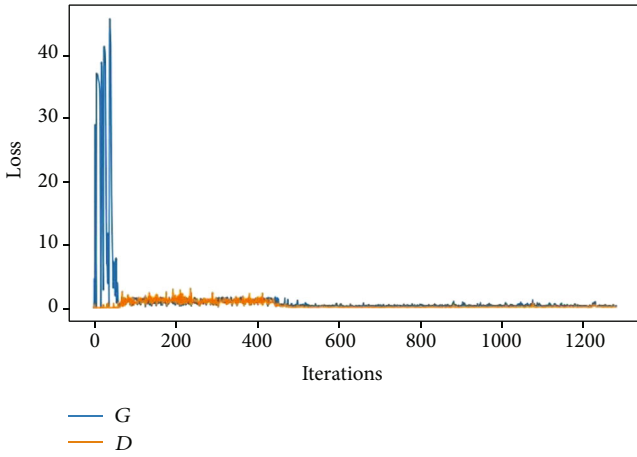


FIGURE 12: Improved DCGAN loss with the mean of initial weight 0.

and the maximum and minimum values of PD become significantly higher. Although the average value of PFA does not significantly decrease, but its maximum and minimum values are closer to the mean, which indicates its fluctuation range becomes smaller with the improved DCGAN.

4.2. *AlexNet*. Figures 15 and 16, respectively, give the PD and PFA comparisons with the AlexNet network under various SNRs, where the improved DCGAN is considered.

From Figure 15, the average value of PD is 0.95 or more, and the minimum value is 0.8 or more when $-2\text{dB} \leq \text{SNR} \leq$

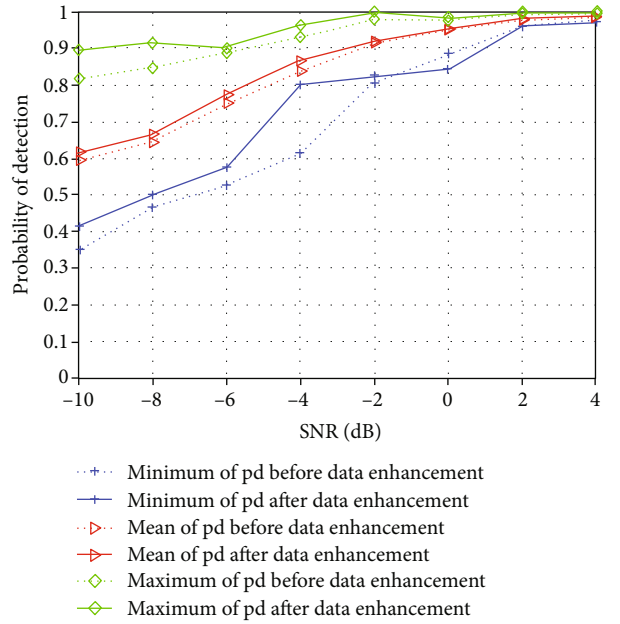


FIGURE 13: Detection probability with LeNet.

4dB. The average value of PD is close to the maximum, which means that most PD values are close to 1, and only a few are close to the minimum values. In addition, after the data enhancement, the average value of PD is also improved slightly.

The PFA is provided with the AlexNet network in Figure 16. From Figure 16, the mean value of PFA is less than 0.05 and its curve almost coincides with the minimum curve. At the same time, after the data enhancement, the maximum value of PFA decreases obviously, which indicates that the vibration amplitude of PFA decreases obviously.

As a summary, when $-2\text{dB} \leq \text{SNR} \leq 4\text{dB}$, PD is close to 1 and PFA is close to 0, which indicates that the detection performance of the AlexNet is better than that of the LeNet. After the data enhancement, PD is improved slightly, and PFA has become more stable.

4.3. *VGG-16*. In Figures 17 and 18, the PD and PFA are, respectively, exhibited with the VGG-16 network. According

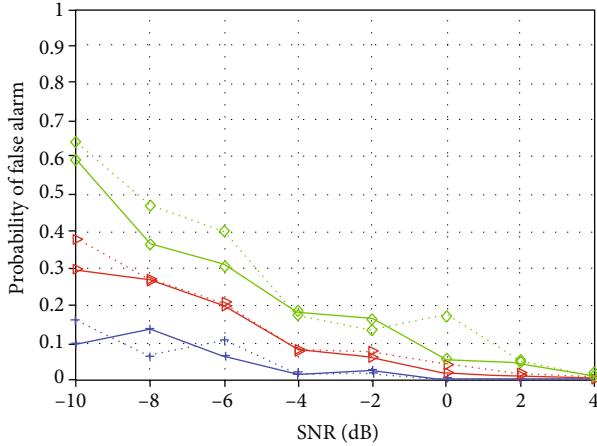


FIGURE 14: False alarm probability with LeNet.

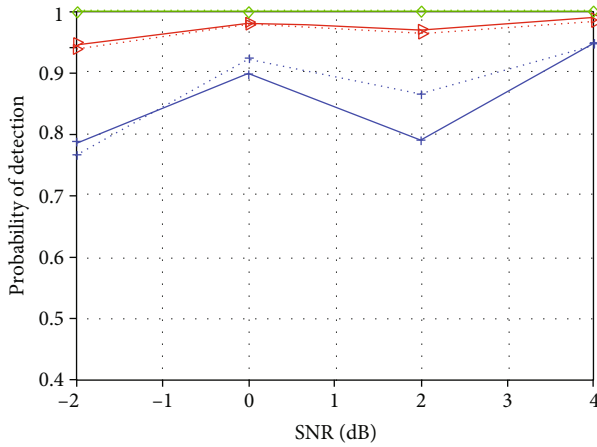


FIGURE 15: Detection probability with AlexNet.

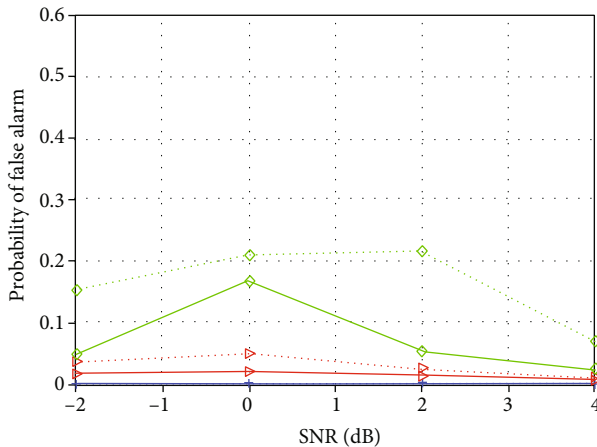


FIGURE 16: False alarm probability with AlexNet.

to Figure 17, when $\text{SNR} = -10$ dB or $\text{SNR} = -4$ dB, the average value of PD is improved by nearly 0.1. When $\text{SNR} \geq -4$ dB, the average value of PD is above 0.9. After the data enhancement, the minimum value of PD is also significantly improved.

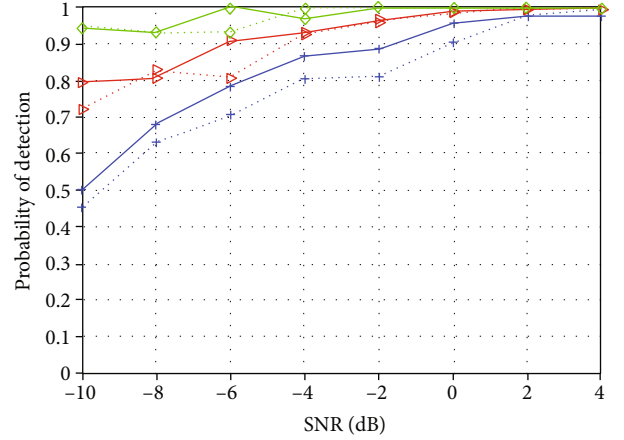


FIGURE 17: Detection probability with VGG-16.

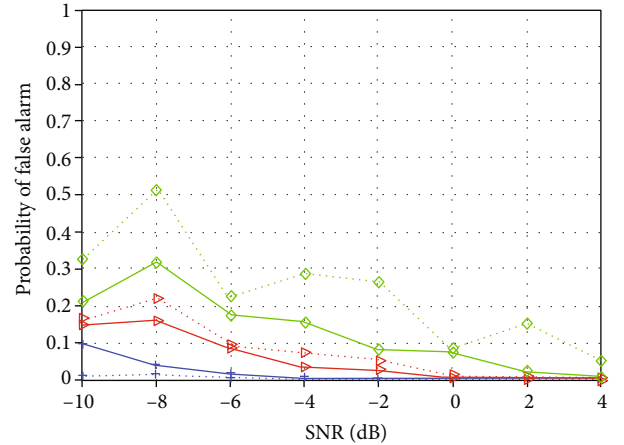


FIGURE 18: False alarm probability with VGG-16.

From Figure 18, the average value of PFA is less than 0.2 and decreases with the increase of SNR. After the data enhancement, the mean and maximum values of PFA decrease significantly.

To sum up, the detection performance of VGG-16 is better than that of LeNet-5 and AlexNet. In addition, the improvement of data enhancement on the performance of the model is also obvious, where the vibration amplitude of PD and PFA decreases obviously. However, the performance of VGG-16 is at the expense of computational complexity due to a large number of network parameters.

4.4. CNN-1. According to the sensing performance of the LeNet, AlexNet, and VGG-16 network, the network depth of VGG-16 is too large and the depth of Lenet-5 is too small to achieve the optimal result. Therefore, this paper designs a novel convolution neural network with the appropriate depth, named CNN-1. The network parameters of the proposed CNN-1 network are shown in Table 2.

As described in the system model, R_x is a real matrix with the dimension $M \times M \times 3$. Let $R_x(i, j, \tau)$ denote the element at position (i, j) of the τ -th dimension for R_x .

TABLE 2: The parameters of CNN-1.

Input: true color image (dimension: $110 \times 110 \times 3$)	
Layers	Kernel size
$C_1 + \text{ReLU}$	$32 \times 3 \times (3 \times 3)$
S_1 : maximum pooling	2×2
$C_2 + \text{ReLU}$	$64 \times 32 \times (3 \times 3)$
S_2 : maximum pooling	2×2
$C_3 + \text{ReLU}$	$128 \times 64 \times (3 \times 3)$
S_3 : maximum pooling	2×2
$C_4 + \text{ReLU}$	$128 \times 128 \times (3 \times 3)$
S_4 : maximum pooling	2×2
$F_1 + \text{ReLU}$	512×3200
$F_2 + \text{sigmoid}$	1×512
Output: score vector (dimension: 1×1)	

- (1) *Convolutional Layer C_1* . C_1 contains 32 feature maps, and each feature map is gained by convolution operation with the kernel size of 3×3 . Thus

$$C_1(i, j, \tau) = \varphi_{\text{ReLU}} \left(\sum_{k=0}^2 \sum_{i_0=0}^2 \sum_{j_0=0}^2 R_x(i + i_0, j + j_0, k) \right) k_{\tau}^{C_1} \cdot (3 - i_0, 3 - j_0, 2 - k), \quad (11)$$

where $C_1(i, j, \tau)$ denotes the element at position (i, j) of the τ -th feature map in C_1 layer, and $k_{\tau}^{C_1}$ denotes the kernel of the τ -th feature map in C_1 layer. $\varphi_{\text{ReLU}} = \max(0, x)$ denotes ReLU function, where $\max(\cdot)$ denotes the maximum value, x is the independent variable of the function.

- (2) *Pooling Layer S_1* . S_1 contains 32 feature maps, and maximum pooling is conducted for each feature map

$$S_1(i, j, \tau) = \max \{ C_1(2i - 1, 2j - 1, \tau), C_1(2i - 1, 2j, \tau), C_1(2i, 2j - 1, \tau), C_1(2i, 2j, \tau) \}. \quad (12)$$

- (3) *Convolutional Layer C_2* . C_2 contains 64 feature maps, and each feature map is gained by convolution operation. Thus

$$C_2(i, j, \tau) = \varphi_{\text{ReLU}} \left(\sum_{k=0}^{31} \sum_{i_0=0}^2 \sum_{j_0=0}^2 R_x(i + i_0, j + j_0, k) \right) k_{\tau}^{C_2} \cdot (3 - i_0, 3 - j_0, 31 - k). \quad (13)$$

- (4) *Pooling Layer S_2* . S_2 contains 64 feature maps, and maximum pooling is conducted for each feature map

$$S_2(i, j, \tau) = \max \{ C_2(2i - 1, 2j - 1, \tau), C_2(2i - 1, 2j, \tau), C_2(2i, 2j - 1, \tau), C_2(2i, 2j, \tau) \}. \quad (14)$$

- (5) *Convolutional Layer C_3* . C_3 contains 128 feature maps, and each feature map is gained by convolution operation

$$C_3(i, j, \tau) = \varphi_{\text{ReLU}} \left(\sum_{k=0}^{127} \sum_{i_0=0}^2 \sum_{j_0=0}^2 R_x(i + i_0, j + j_0, k) \right) k_{\tau}^{C_3} (3 - i_0, 3 - j_0, 127 - k). \quad (15)$$

- (6) *Pooling Layer S_3* . S_3 contains 128 feature maps, and maximum pooling is conducted for each feature map

$$S_3(i, j, \tau) = \max \{ C_3(2i - 1, 2j - 1, \tau), C_3(2i - 1, 2j, \tau), C_3(2i, 2j - 1, \tau), C_3(2i, 2j, \tau) \}. \quad (16)$$

- (7) *Convolutional Layer C_4* . C_4 contains 128 feature maps, and each feature map is gained by convolution operation

$$C_4(i, j, \tau) = \varphi_{\text{ReLU}} \left(\sum_{k=0}^{127} \sum_{i_0=0}^2 \sum_{j_0=0}^2 R_x(i + i_0, j + j_0, k) \right) k_{\tau}^{C_4} (3 - i_0, 3 - j_0, 127 - k), \quad (17)$$

- (8) *Pooling Layer S_4* . S_4 contains 128 feature maps, and each feature map is gained by convolution operation

$$S_4(i, j, \tau) = \max \{ C_4(2i - 1, 2j - 1, \tau), C_4(2i - 1, 2j, \tau), C_4(2i, 2j - 1, \tau), C_4(2i, 2j, \tau) \}. \quad (18)$$

- (9) *Fully Connected Layer F_1* . F_1 is fully connected with S_4 , the number of neurons in F_1 is 512

- (10) *Fully Connected Layer F_2* . F_2 is fully connected with F_1 with the neuron number 512. F_2 is a probability value because SS is a binary classification issue. This probability value can be expressed as $P_{L_1}(R_x)$. When $P_{L_1}(R_x) > 0.5$, $R_x \in H_1$; otherwise, $R_x \in H_0$.

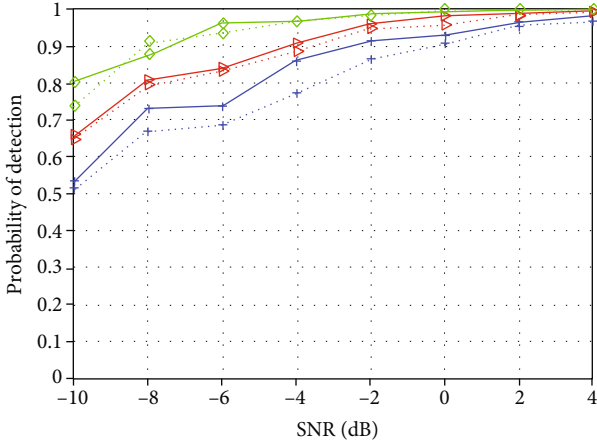


FIGURE 19: Detection probability with CNN-1.

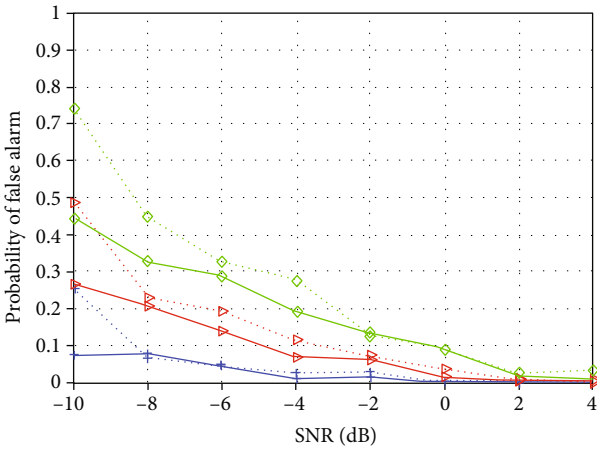


FIGURE 20: False alarm probability with CNN-1.

Figures 19 and 20, respectively, exhibit the PD and PFA with the proposed CNN-1 scheme under various SNRs. From Figures 19 and 20, the detection probability of the CNN-1 with the proposed data enhancement scheme is higher than that without the data enhancement while the false alarm probability of the CNN-1 with the proposed data enhancement scheme is lower than that without the data enhancement. In addition, when $\text{SNR} \geq -8\text{dB}$, the mean value of PD is more than 0.8 and the mean value of PFA is less than 0.2, which is much better than that of the LeNet and the same as that of the VGG-16.

Figure 21 gives the performance comparisons of various SS schemes including the LeNet based scheme, the AlexNet based scheme, the VGG-16 based scheme, and the proposed CNN-1 based scheme. As it is shown for the black line, it, respectively, denotes -2 dB, 0 dB, 2 dB, and 4 dB from bottom to top, and the other lines are the same as the black one. From Figure 21, the sensing performance of the CNN-1 based SS scheme is highest compared to the other SS schemes under the same SNR, which corresponds with the performance in Figures 13–20.

In Figure 22, the computation time comparison of different CNN networks is discussed, where we use the time it

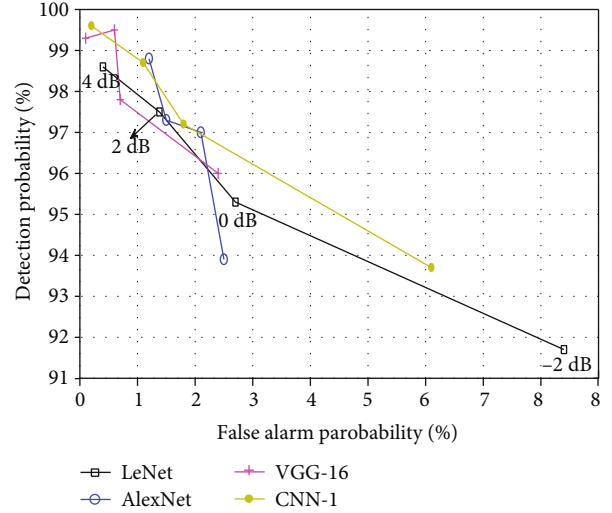


FIGURE 21: Performance comparisons of various SS schemes.

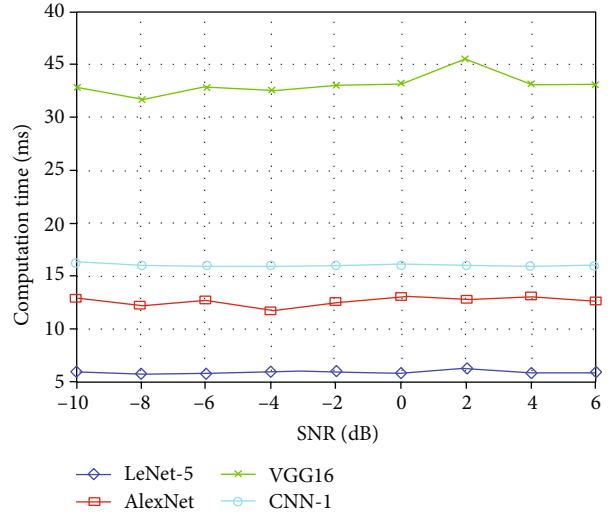


FIGURE 22: Computation time comparisons of different CNN networks.

takes for a model to process a true color image to represent computational complexity and evaluate the computational complexity of different CNN algorithms. From Figure 22, the computation time of the LeNet, AlexNet, CNN-1, and VGG-16 is about 6 ms, 12.5 ms, 16 ms, and 33 ms, which indicates that the LeNet is the simplest and the VGG-16 is the most complex.

As a summary, the detection performance of CNN-1 is similar to that of VGG-16, but the computation time of CNN-1 is nearly half that of VGG-16. This indicates that CNN-1 is a better convolution neural network model for SS.

As a supplement, performance comparisons are made between the proposed CNN-1 and the schemes in [16, 17]. As it is shown for the black line, it, respectively, denotes -2 dB, 0 dB, 2 dB, and 4 dB from bottom to top, and the other lines are the same as the black one. From Figure 23, the sensing performance of the proposed CNN-1 scheme is slightly higher than that of the scheme in [16]. Meanwhile, the

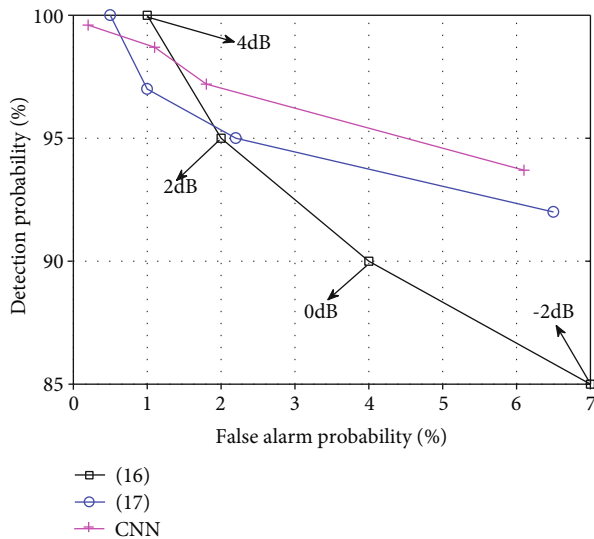


FIGURE 23: Performance comparisons between CNN-1 with the schemes in [16, 17].

sensing performance of the proposed CNN-1 scheme is significantly higher than that of the scheme in [17]. This states that the proposed CNN-1 scheme with data enhancement is more suitable for the detection of spectrum state.

5. Conclusions

In this paper, the deep learning based spectrum sensing is discussed for sustainable cities and society, where the LeNet, AlexNet, VGG-16, and the proposed CNN-1 network are considered. First, the two-dimensional dataset of the received signal is established and expanded by the proposed DCGAN scheme. Then, the four networks are trained on the expanded dataset. Finally, the test is made under the various SNR conditions. The experiment results show that the sensing performance is greatly improved by the proposed data enhancement scheme and the novel CNN network.

Data Availability

The data supporting the results of my study can be found at https://figshare.com/articles/dataset/___part1_rar/14245763.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] Y. Wu, B. Rong, K. Salehian, and G. Gagnon, "Cloud transmission: a new spectrum-reuse friendly digital terrestrial broadcasting transmission system," *IEEE Transactions on Broadcasting*, vol. 58, no. 3, pp. 329–337, 2012.
- [2] N. Chen, B. Rong, X. Zhang, and M. Kadoch, "Scalable and flexible massive MIMO precoding for 5G H-CRAN," *IEEE Wireless Communications*, vol. 24, no. 1, pp. 46–52, 2017.
- [3] Y. Cui, F. Liu, X. Jing, and J. Mu, "Integrating sensing and communications for ubiquitous IoT: applications, trends and challenges," 2021, <http://arxiv.org/abs/2104.11457>.
- [4] A. Bishnu and V. Bhatia, "LogDet covariance based spectrum sensing under colored noise," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 7, pp. 6716–6720, 2018.
- [5] K. Zheng, X.-Y. Liu, X. Liu, and Y. Zhu, "Hybrid overlay-underlay cognitive radio networks with energy harvesting," *IEEE Transactions on Communications*, vol. 67, no. 7, pp. 4669–4682, 2019.
- [6] J. I. Mitola and G. Q. Maguire, "Cognitive radio: making software radios more personal," *IEEE Personal Communications*, vol. 6, no. 4, pp. 13–18, 1999.
- [7] K. Yadav, S. Dhar Roy, and S. Kundu, "Throughput of cognitive radio networks with improved energy detector under security threats," *International Journal of Communication Systems*, vol. 31, no. 6, article e3512, 2018.
- [8] Y. Lu, "A sensing contribution-based two-layer game for channel selection and spectrum access in cognitive radio ad-hoc networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 6, pp. 3631–3640, 2018.
- [9] E. Ollila and E. Raninen, "Optimal shrinkage covariance matrix estimation under random sampling from elliptical distributions," *IEEE Transactions on Signal Processing*, vol. 67, no. 10, pp. 2707–2719, 2019.
- [10] X. Hao, G. Zhang, and S. Ma, "Deep learning," *International Journal of Semantic Computing*, vol. 10, no. 3, pp. 417–439, 2016.
- [11] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning based semantic communications: an initial investigation," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, Taipei, Taiwan, December 2020.
- [12] S. Zheng, S. Chen, and X. Yang, "DeepReceiver: a deep learning-based intelligent receiver for wireless communications in the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 7, no. 1, pp. 5–20, 2021.
- [13] C. Luo, J. Ji, Q. Wang, X. Chen, and P. Li, "Channel state information prediction for 5G wireless communications: a deep learning approach," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 1, pp. 227–236, 2020.
- [14] Q. Cheng, Z. Shi, D. N. Nguyen, and E. Dutkiewicz, "Non-cooperative OFDM spectrum sensing using deep learning," in *2020 International Conference on Computing, Networking and Communications (ICNC)*, Big Island, HI, USA, February 2020.
- [15] J. Gao, X. Yi, C. Zhong, X. Chen, and Z. Zhang, "Deep learning for spectrum sensing," *IEEE Wireless Communications Letters*, vol. 8, no. 6, pp. 1727–1730, 2019.
- [16] S. Zheng, S. Chen, P. Qi, H. Zhou, and X. Yang, "Spectrum sensing based on deep learning classification for cognitive radios," *China Communications*, vol. 17, no. 2, pp. 138–148, 2020.
- [17] J. Xie, J. Fang, C. Liu, and X. Li, "Deep learning-based spectrum sensing in cognitive radio: a CNN-LSTM approach," *IEEE Communications Letters*, vol. 24, no. 10, pp. 2196–2200, 2020.
- [18] S. Lin, L. Cai, X. Lin, and R. Ji, "Masked face detection via a modified LeNet," *Neurocomputing*, vol. 218, pp. 197–202, 2016.
- [19] M. Alencastre-Miranda, R. R. Johnson, and H. I. Krebs, "Convolutional neural networks and transfer learning for quality

- inspection of different sugarcane varieties,” *IEEE Transactions on Industrial Informatics*, vol. 17, no. 2, pp. 787–794, 2021.
- [20] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *Computer Science*, 2014.
- [21] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” *Computer Science*, 2015.
- [22] J. Zhang, L. Chen, L. Zhuo, X. Liang, and J. Li, “An efficient hyperspectral image retrieval method: deep spectral-spatial feature extraction with DCGAN and dimensionality reduction using t-SNE-based NM hashing,” *Remote Sensing*, vol. 10, no. 2, p. 271, 2018.
- [23] Y. Liu, J. Yi, X. Wan, X. Zhang, and H. Ke, “Time varying clutter suppression in CP OFDM based passive radar for slowly moving targets detection,” *IEEE Sensors Journal*, vol. 10, no. 2, p. 271, 2020.
- [24] M. Ranjbar, N. H. Tran, M. Vu, T. V. Nguyen, and M. Cenk Gursoy, “Capacity region and capacity-achieving signaling schemes for 1-bit ADC multiple access channels in Rayleigh fading,” *IEEE Transactions on Wireless Communications*, vol. 19, no. 9, pp. 6162–6178, 2020.
- [25] R. Umar, A. U. H. Sheikh, and M. Deriche, “Unveiling the hidden assumptions of energy detector based spectrum sensing for cognitive radios,” *IEEE Communications Surveys & Tutorials*, vol. 16, no. 2, pp. 713–728, 2014.
- [26] J. Mu, X. Jing, H. Huang, and N. Gao, “Subspace based method for spectrum sensing with multiple users over fading channel,” *IEEE Communications Letters*, vol. 22, no. 4, pp. 848–851, 2018.

Research Article

Influencing Factors and Forecasting Statistics of Enterprise Market Sales Based on Big Data and Intelligent IoT

Zhen Guo¹ and Tao Zou² 

¹Business School, University of Queensland, Brisbane Qld 4072, Australia

²School of Management, Northwestern Polytechnic University, Xi'an, 710000 Shaanxi, China

Correspondence should be addressed to Tao Zou; zoutolele@mail.nwpu.edu.cn

Received 2 April 2021; Revised 27 May 2021; Accepted 25 June 2021; Published 26 July 2021

Academic Editor: Bo Rong

Copyright © 2021 Zhen Guo and Tao Zou. 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 acceleration of economic development, enterprise management is facing more severe challenges. Big data analysis based on the intelligent Internet of Things (IoT) has a positive effect on the development of enterprise management and can make up for the shortcomings of enterprise management. In this paper, we develop a big data processing method based on intelligent IoT which can mine the factors that affect the company's market sales from the collected data. Then, we propose a KNN classification algorithm based on overlapping k -means clustering. This algorithm adds a training process to the traditional KNN algorithm, which can accurately classify data and greatly improve the efficiency of the classification algorithm. Numerical analysis results prove the effectiveness of the proposed algorithm.

1. Introduction

In recent years, with the rapid development of social economy and science and technology, big data analysis in Intelligent Internet of Things (IoT) has been widely applied to various industries and fields. It can help enterprises find problems existing in management and promote the improvement of enterprise management level. Applying big data analysis in intelligent IoT to enterprise management can adapt to the changes of internet enterprises, help enterprises better cope with various challenges, and lay a good foundation for the sustainable development of enterprises.

The development of modern science and technology also drives the development of intelligent IoT industry. Intelligent IoT is a concept that emerged in 2018. It refers to the system collects all kinds of information in real time through various information sensors (generally in the context of monitoring, interaction, and connection) and makes intelligent analysis of data through machine learning in terminal devices, edge domains, or cloud centers, including positioning, comparison, prediction, and scheduling. At the technical level, AI enables the Internet of Things to acquire the perception

and recognition ability, and the Internet of Things provides the data of the training algorithm for AI.

As big data applications continue to penetrate into all walks of life around the world and take root, traditional data management methods no longer meet the data management needs of enterprises [1]. In the future, the surrounding environment of enterprises in business activities is unpredictable. To increase the risks and opportunities in its business activities, if it can predict the sales volume in the business activities, use big data technology to analyze the factors influencing the company's market sales, and formulate response strategies in advance, it can be better at resisting risks and transformation opportunities will ultimately increase the company's profit in the market and stabilize its leading position in business activities.

When applying big data technology in intelligent IoT to enterprise management, a large number of real data of enterprises can be extracted and processed and analyzed with big data technology, so as to provide reliable reference basis. Compared with the traditional data management system, the system architecture of the Internet of Things includes the following parts:

- (1) The LAAS layer, which is the important data storage layer of the Internet of Things system, can select cloud for data storage to facilitate data query and utilization
- (2) PaaS layer, mainly to provide the development languages and tools required by customers, such as Python, Hive, and Hadoop
- (3) SaaS layer, mainly to provide the applications needed by customers, to facilitate the use of devices for client interface access, such as intelligent large screen, PC terminal, and common client interface. The data information of each business and each sales market can be monitored. Due to the different requirements for data information in data management of various enterprises, the application of big data analysis in intelligent IoT should be combined with its own situation, and the systematic analysis of the existing data resources of the enterprise should be carried out to help the enterprise find its own problems and find the best solution

Many foreign scholars have conducted research on the influencing factors of corporate market sales and forecast statistics and achieved good results. For example, Chemmanur et al. designed an attribute cleaning method for different categories of “dirty data” and proposed a method based on the tree-structured Bloom Filter algorithm cleans duplicate data. Massive data has been cleaned by multiple iterations to ensure the data quality of the following sales forecast analysis [2]. Singh and Mohanty proposed a new type of online data storage and processing model-online analytical processing. In terms of data analysis and processing functions, OLAP has greatly improved compared with OLTP and can meet various application needs of users [3]. Yadegaridehkordi et al. proposed a combined forecasting model, that is, analyzing the characteristics of a single forecasting model, and then combining different models according to a certain weight ratio, so as to give full play to the advantages of different models and improve the prediction accuracy of the model [4].

In the research of related scholars in our country, Wang and Han designed an extended radial basis function as the kernel function for the multidimensional and nonlinear characteristics of the product sales sequence and used an improved immune optimization algorithm to adjust the parameters carry out optimization, establish a set of support vector machine forecasting system, apply the system to car sales forecasting examples, and verify the feasibility of the system by comparing with BP neural network and general radial basis function forecasting accuracy [5]. Gu et al. use a background value optimization formula with adjustment factors to optimize the gray GM model and applies the established optimization model to the number of tourists and tourism income in Hangzhou’s tourism industry to predict [6]. Xin et al. provide guidance and suggestions for the tourism industry to formulate sales strategies [7]. Xing et al. use principal component analysis to reduce the dimensions of 8 related clothing promotion factors, use particle swarm optimization algorithm to optimize the neural network, establish

a clothing product sales forecast model, reduce the training time of the network, and improve the accuracy of the network prediction degree [8].

Facing a large amount of enterprise data acquisition, storage, and analysis work, we use big data technology in intelligent IoT to ease the workload of data analysts. This paper uses the big data analysis method in intelligent IoT to analyze the historical sales data of enterprises, and its time series presents the characteristics of dual trend changes. According to its characteristics, a combination forecasting model based on the trend method and seasonal index method is proposed, and MAE, RMSE, and MAPE forecast evaluation standards are used to forecast the combination the model is compared with the forecast effect of a single-trend forecast model and seasonal index model, and then, the optimal forecast model is selected.

2. Influencing Factors and Forecasting Statistical Model of Enterprise Market Sales Based on Big Data Analysis in Intelligent IoT

2.1. Intelligent IoT Architecture. It mainly includes three levels: intelligent devices and solutions, the OS layer, and infrastructure in intelligent IoT architecture, which are finally delivered through integration services, as shown in Figure 1. Intelligent equipment can achieve the data collection of view, audio, and pressure and perform the action of capturing, sorting, and handing. Usually IoT devices and solutions are provided to customers. This layer involves diversification of device forms. The OS layer is equivalent to the “brain” of intelligent IoT, which can mainly connect and control the device layer, provide intelligent analysis and data-handling capacity, and solidify the core applications for scenarios into function modules. This layer has high requirements on business logic, unified modeling, full-link technical capacity, and high-concurrency supporting capacity. The infrastructure layer provides the IT infrastructure of servers, storage, AI training, and deployment capabilities, etc. In the era of intelligent IoT, mass data generated in the production and life of people will be collected by sensors in intelligent IoT. In the era of big data, the individual behavior of consumers not only can be collected, quantified, and predicted but also consumers’ personal opinions may change the operation of the business society.

2.2. Use MapReduce to Simplify Massive Data. MapReduce adopts the master-slave mode, which is to set up a master node and multiple slave nodes to jointly complete the entire process of distributed computing [9, 10]. In the calculation process, data processing needs to be carried out through the cooperation of two stages of map (mapping function) and reduce (reduction function). Generally, the output of one stage is the input of the next stage, and the two require multiple coordination and cooperation.

2.2.1. JobTracker. JobTracker is mainly used to receive the application processing program, resource monitoring, and job scheduling submitted by the client. This part of the program mainly depends on the programmer to make some

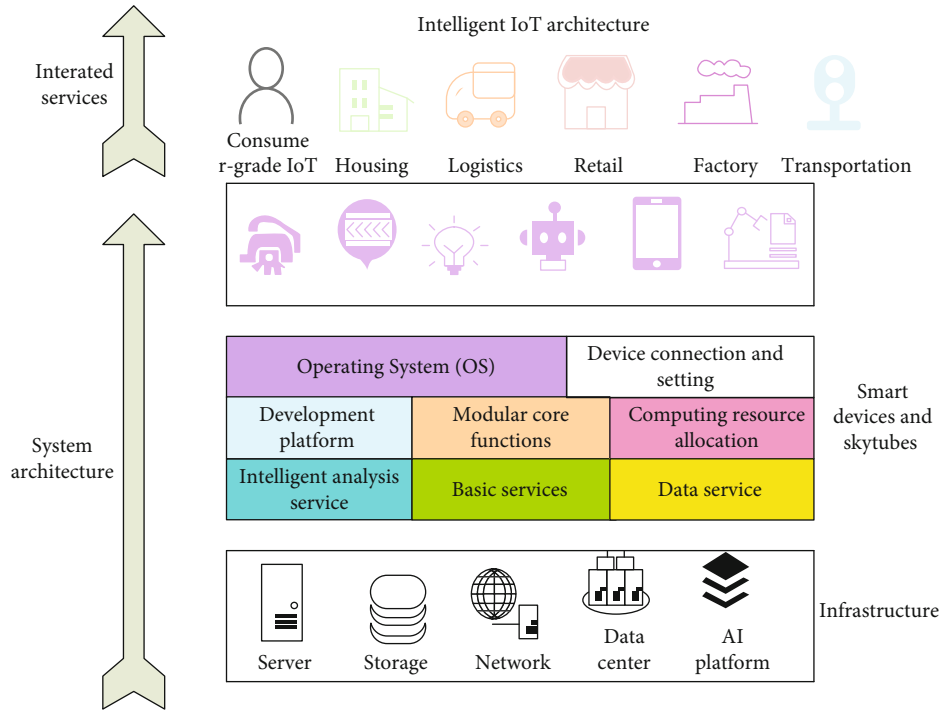


FIGURE 1: Intelligent IoT architecture.

complex algorithms. It monitors the running status of TaskTracker and jobs in the system by accepting the heartbeat information sent by TaskTracker. The good fault tolerance mechanism of JobTracker enables TaskTracker or jobs to run abnormally, and tasks running on TaskTracker can be aborted due to abnormalities the task is backed up and executed on other TaskTracker to ensure the stability and reliability of the system [11, 12].

2.2.2. *Client.* The client is mainly a layer where the user interacts with the MapReduce process. This layer mainly plays the role of input, passing the program that the user needs to perform operations to the JobTracker. During the operation, the operation process can be monitored through the console.

2.2.3. *TaskTracker.* The task scheduler is mainly used to coordinate specific calculations and responds to the JobTracker with information such as the time spent in the calculation process, the number of processing tasks, the occupied CPU, and memory, and at the same time, it processes the assigned tasks [13, 14]. Map and reduce are coordinated to complete. MapReduce uses split as the smallest processing unit of data and is used to store the corresponding data block to be processed. Each split will be processed by the corresponding Map Task.

2.3. *Introduction to Machine Learning Classification Algorithms.* Machine learning is essentially an approximation to the real model of the problem. Among them, the supervised classification algorithm has been widely used in many business scenarios. There are many ways to solve classifica-

tion problems. The basic classification methods mainly include decision trees [15], naive Bayes [16], support vector machines [17], K -nearest neighbors [18], and artificial neural networks [19].

The decision tree algorithm is a method of approximating the value of a discrete function. It is a typical classification method. It first processes the data, uses induction algorithms to generate readable rules and decision trees, and then uses decisions to analyze new data. Naive Bayes classification is a classification method based on Bayes' theorem and the assumption of the independence of characteristic conditions. It originated from classical mathematical theory and has a stable mathematical foundation and classification efficiency. A support vector machine is a supervised learning method, which can be widely used in statistical classification and regression analysis. K -nearest neighbor algorithm, referred to as KNN (k -nearest neighbor), is also a relatively simple classification and prediction algorithm. For selecting the K training data that are most similar to the data to be classified and predicted, the results of the K data or the classification labels are averaged and the mode is taken to obtain the results or classification labels of the data to be classified and predicted. The artificial neural network, abbreviated as neural network or quasineural network, is a mathematical model or calculation model that imitates the structure and function of biological neural network and is used to estimate or approximate functions. The neural network is calculated by connecting a large number of artificial neurons. In most cases, the artificial neural network can change the internal structure on the basis of external information and is an adaptive system.

2.4. KNN Classification Algorithm Improved KNN Data Mining Algorithm and Time Series Prediction Algorithm Research

2.4.1. *KNN Algorithm Classification.* Select the K samples with the smallest distance from the sample to be classified as the K -nearest neighbors of X and finally judge the category of X based on the K -nearest neighbors of X .

(1) *Algorithm Flow.* Calculate the distance between the sample to be classified and each training sample: the distance function in the KNN algorithm generally has Euclidean distance:

$$d_{\text{euc}}(x, y) = \left[\sum_{j=1}^d (x_j - y_j)^2 \right]^{1/2} = [(x - y)(x - y)^T]^{1/2},$$

$$d_{\text{mah}}(x, y) = \sqrt{(x - y) \sum (x - y)^T}. \quad (1)$$

Manhattan distance:

$$d_{\text{mah}}(x - y) = \sum_{j=1}^d |x_j - y_j|. \quad (2)$$

Chebyshev distance:

$$d_{\text{che}}(x, y) = \max_j (|x_j - y_j|). \quad (3)$$

As well as Min's distance, average distance, and geodetic distance, among these distances, the Euclidean distance is often used because of its simplicity.

(2) *Selection of Prediction Algorithm.* Because the sales of cigarettes need to be predicted, and the sales data of cigarettes is not a continuous time, but a time point, it is more appropriate to choose a time series model when choosing a forecasting algorithm, and the time series can establish a relationship that includes dynamic dependencies. Based on the data model, the trend of future data can be observed from historical behavior information [20, 21].

2.4.2. *Improve Data Mining Algorithm.* This paper introduces the training process of clustering after partitioning into the KNN algorithm. That is, the big data is divided into equidistant blocks, and then, the data is clustered on each block of data. In this way, for big data, dividing the data into many blocks can effectively reduce the requirements for computer memory, so that the KNN classification algorithm can be used in a big data environment.

First, the big data is divided into blocks. According to different requirements, it can be divided into blocks according to different blocking methods. Given n data samples $\{x_1, x_2, \dots, x_n\}$, find K clusters Class center $\{a_1, a_2, \dots, a_n\}$, so that the sum of squared distances between each data sample and its nearest cluster center is the smallest. This sum of squared

TABLE 1: Demographic factors and the trend of cigarette sales.

	Permanent residents	Floating population	Total employees	Product sales	Sales amount
2016	573	66	383	200	86
2017	615	69	416	248	235
2018	689	104	447	307	462
2019	805	497	601	408	863

distances is called the objective function W_N , and its mathematical expression is formula (4), the data is recorded as data matrix formula (5), the difference between data and data uses dissimilarity matrix formula (6), according to the characteristics of the data in this article, the objective function W_N is transformed into formula (7):

$$W_n = \sum_{i=1}^n \min_{1 \leq j \leq k} |x_i - a_j|^2, \quad (4)$$

$$\begin{bmatrix} x_{11} & \cdots & x_{1j} & \cdots & x_{1p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{i1} & \cdots & x_{ij} & \cdots & x_{ip} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{n1} & \cdots & x_{nj} & \cdots & x_{np} \end{bmatrix}, \quad (5)$$

$$\begin{bmatrix} 0 & & & & \\ & d(2, 1) & 0 & & \\ & d(3, 1) & d(3, 2) & 0 & \\ & \cdots & \cdots & \cdots & 0 \\ d(n, 1) & d(n, 2) & \cdots & \cdots & 0 \end{bmatrix}, \quad (6)$$

$$W_N \sum_{i=1}^n \min_{1 \leq j \leq k} \left| \sqrt{(x_{ik} - a_{jk})^2} \right|^2. \quad (7)$$

In formula (5), each column represents a data attribute, and each row represents a piece of data. In formula (6), $d(m, n)$ represents the degree of difference between the m th data and the n th data. When the difference between the two data is smaller, the value of $d(m, n)$ will also be smaller [22, 23].

2.4.3. *Time Series Stationarity Test.* According to the calculation formula of unit root in MyEclipse, if the sequence is nonstationary, there is $\beta = 1$, and if the sequence is stationary, then $\beta < 1$. Now, assuming that the sequence is nonstationary, let $\beta = 11$ into formula (8), (9) and calculate the value of DF:

$$s(\hat{\beta}) = \frac{\sqrt{(1/T - 1) \sum_{t=2}^T \hat{u}_t^2}}{\sqrt{\sum_{t=2}^T y_{t-1}^2}}, \quad (8)$$

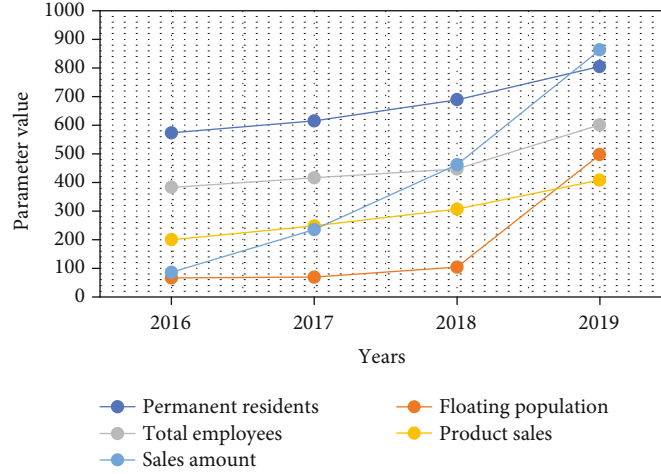


FIGURE 2: Demographic factors and the trend of cigarette sales.

$$DF(X) = \frac{\hat{\beta} - 1}{s(\hat{\beta})}. \quad (9)$$

The running result is the unit root test value of the original data. According to the value, it can be judged whether the original data is stable. When the value does not exist or is particularly small, the original data can be judged to be a stationary sequence; otherwise, it is not stable [24, 25].

2.4.4. Time Series Smoothing Processing. Because the original sequence is a nonstationary sequence, the difference calculation is performed on the original sequence for stationary processing. The calculation formula is

$$\begin{aligned} \Delta x_t &= x_t - x_{t-1}, \\ \Delta^2 x_t &= \Delta x_t - \Delta x_{t-1}, \\ \Delta^d x_t &= \Delta^{d-1} x_t - \Delta^{d-1} x_{t-1}, \end{aligned} \quad (10)$$

Among them, t is the time point. If there are periodic fluctuations in the time series, then the data should also be subjected to a seasonal difference operation. The seasonal difference processing operation can clear the periodicity of the time series. The calculation formula is

$$\Delta_s X_t = X_t - X_{t-s}. \quad (11)$$

2.5. Second Moving Average Method. Establishing a forecasting model for the second moving average method is the key to forecasting using this method. The forecasting model of the second moving average forecasting method is shown in formula (12):

$$Y_{t+1} = a_t + b_t \times T, \quad (12)$$

$$a_t = 2M_t^{(1)} - M_t^{(2)}, \quad (13)$$

$$b_t = \frac{2}{n-1} (M_t^{(1)} - M_t^{(2)}). \quad (14)$$

The T in the above formula represents the expectation that starts at time t and moves backward. $M_t^{(1)}$ is the last moving average in the first moving average sequence obtained by calculation. $M_t^{(2)}$ represents the last moving average in the second moving average sequence [26, 27]. Correspondingly, the formulas for calculating the primary and secondary moving average are as follows:

$$M_t^{(1)} = \frac{Y_t + Y_{t-1} + \dots + Y_{t-n+1}}{n}, \quad (15)$$

$$M_t^{(2)} = \frac{M_t^{(1)} + M_{t-1}^{(1)} + \dots + M_{t-n+1}^{(1)}}{n}. \quad (16)$$

Formulas (15) and (16) represent the time series observation values to be predicted, and $M_t^{(1)}$ and $M_t^{(2)}$ represent the primary and secondary moving average values of period t , respectively and n is the spanning dimension of this calculation. The basic prediction formula of the exponential smoothing prediction model is

$$S_t = 2Y_t + (1-a)S_{t-1}. \quad (17)$$

In formula (17), when S_{t-1} represents time t , the actual value Y_{t-1} at that time corresponds to the smooth value, and S_{t-1} represents the smooth value corresponding to the actual value Y_{t-1} at time $t-1$. The parameter a in the formula is a weight value, which is also called a smoothing constant under normal circumstances, and the value range is $[0,1]$.

2.6. Design of Enterprise Marketing System Based on Hadoop

2.6.1. System Architecture. The design of the enterprise marketing system is to build a Hadoop-based data processing platform as a data management center and provide massive data storage and processing support to implement a Hadoop-based enterprise marketing system [28, 29].

(1) Data Source Layer. The main job of the data source layer is to collect data. The data source of the enterprise marketing

TABLE 2: Personal spending power and the trend of cigarette sales.

	Per capita consumption amount	GDP per capita	Per capita disposable expenditure	Per capita consumption expenditure	Average salary
2016	737	7215	3665	2654	7976
2017	1654	7567	4317	3475	8242
2018	1964	7988	4784	3864	8527
2019	2268	8365	5521	4269	8731

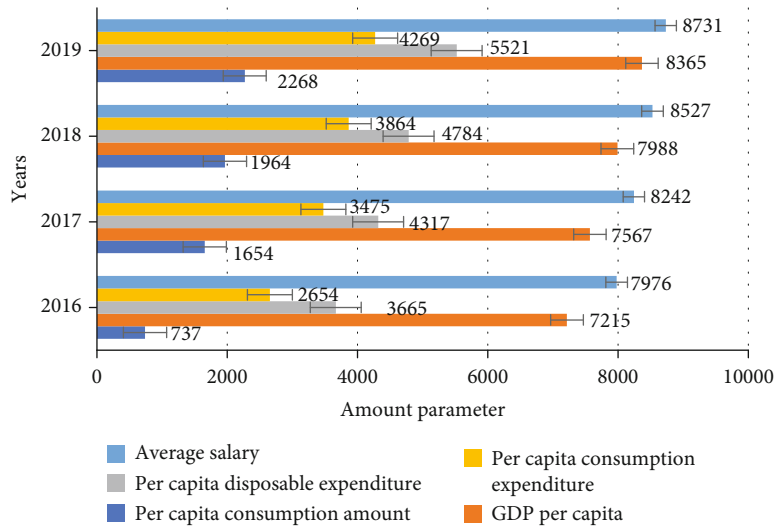


FIGURE 3: Personal spending power and the trend of cigarette sales.

system in this article includes internal data and external data. The data collection method is mainly through the National Bureau No. 1 Project. Downstream data includes the production and sales information of more than a dozen industrial companies across the country, as well as the company's purchase, sales, and inventory data in various markets across the country. Salespersons report market data and import external systems through the Web Services interface.

(2) *Data Transmission Layer*. There are two main ways of data transmission; one is through ETL middleware, and the other is a data transmission interface through enterprise applications. Generally, small-scale data provides a data transmission interface, and the data can be transmitted directly through the interface, while for large-scale data extraction, data is directly extracted by connecting to the database through ETL middleware.

(3) *Data Processing Layer*. Since the source data transmitted from the data source layer is not only the finest granularity but also the amount of data is very large, and there are many "dirty data," the transmitted data needs to be preprocessed before storage including data cleaning and processing. Due to the huge amount of data, data processing is performed by the Hadoop platform to achieve dimensionality reduction and aggregation of massive data and simplify the data on the basis of satisfying model analysis and maintaining data integrity and accuracy.

TABLE 3: Comparison of sales forecast results and actual values based on trend model.

Month	Forecast sales	Guess sales	Actual sales
201801	528739	562284	572639
201802	217479	318374	302746
201803	337468	337595	347290
201804	352648	347284	342699
201805	313947	359837	368270
201806	301748	364829	352890

(4) *Data Storage Center*. The data storage center of the system in this paper is coordinated by the Hadoop distributed storage platform and the relational database. The Hadoop platform is built to store massive data, process massive data, and transmit data to the system [30]. After the data is exported from the Hadoop platform to the relational database, the relational database performs real-time analysis or mining on it to ensure data consistency during processing. Using Hadoop and relational databases to work together to process computing tasks, separate the processing of massive data from the processing of real-time data, so that large-scale data operations will not affect the operating efficiency of the marketing system, and also make the entire system easier for expansion, more stable.

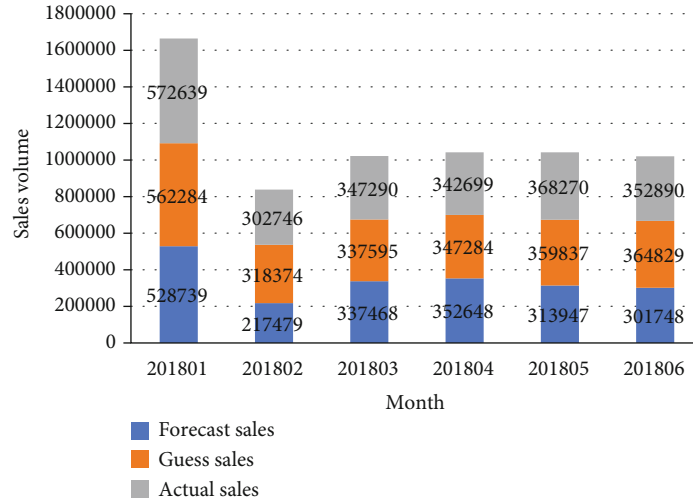


FIGURE 4: Comparison of sales forecast results and actual values based on trend model.

(5) *Data Analysis Layer*. The data analysis logic is mainly realized through the stored procedures of the relational database and the MapReduce calculation model. Since the data sets of small data scale are designed to be stored in the relational database, the analysis of this type of data can be realized through the stored procedures of the relational database. Large-scale data sets are stored in HDFS, and the data model needs to be implemented by writing a MapReduce calculation model.

2.6.2. System Technical Architecture. This paper uses the J2EE enterprise-level application development framework and adopts the stable and convenient B/S operation mode and component development technology to design and develop the system architecture [31]. Data transmission uses XML format as the data transmission standard for each interface and realizes the application integration and data integration of each system through message middleware, and realizes the integration of data collection, data storage, and data pre-processing. Building a Hadoop-based data processing platform as the system's data management center [32], which is low-coupled with other running hardware devices, processes, and stores massive amounts of data at high speed, and can allocate computing and storage resources to other running systems, and use the Hadoop platform working with relational data, it separates the processing of massive data from business logic and analysis operations, reduces the coupling degree of analysis and calculation at the hardware level, and greatly improves the analysis and calculation performance and stability of the system.

3. Influencing Factors and Forecasting Statistical Experiments of Enterprise Market Sales Based on Big Data Analysis in Intelligent IoT

3.1. Experimental Subjects and Data Collection. This article selects the monthly sales volume of a certain brand of ciga-

TABLE 4: Comparison of sales forecast results and actual values based on seasonal index method.

Month	Forecast sales	Guess sales	Actual sales
201801	538478	562284	572639
201802	286491	318374	302746
201803	361937	337595	347290
201804	358505	347284	342699
201805	347190	359837	368270
201806	365281	364829	352890

rettes from China Tobacco in this province from 2016 to 2019 as the analysis data set. The time series of the brand's sales volume has a strong upward trend and also has periodic phenomena, so the time series has the characteristics of dual trend changes, namely, seasonal volatility and overall trend variability, the desired effect will not be achieved. In this paper, the single-term model of the trend method and the seasonal index method will be fitted to forecast, respectively, and then, the linear combination forecasting model of the two will be established to forecast the sales, and finally the forecast results will be compared.

3.2. Forecast Method

- (1) Seasonal index method the calculation method and steps are as follows:
 - (a) Calculate the average of the same quarter over the years. Suppose the average of the same quarter over the years is r_i , $i = 1, 2, 3, 4$. A total of 12 quarterly time series in three years are represented as y_1, y_2, \dots, y_{12} , then you can get:

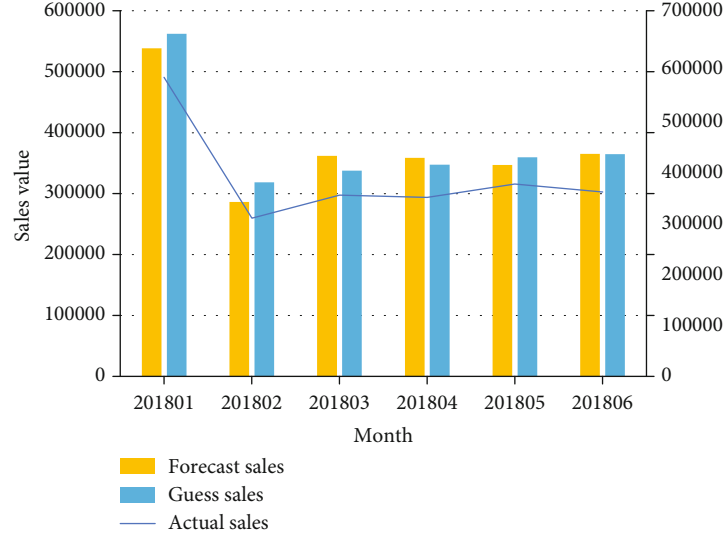


FIGURE 5: Comparison of sales forecast results and actual values based on seasonal index method.

$$r_1 = \frac{1}{n}(y_1 + y_5 + \dots + y_{4n-3}),$$

$$\dots$$

$$r_4 = \frac{1}{n}(y_4 + y_8 + \dots + y_{4n})$$
(18)

(b) Calculate the average of each season. Let \bar{y}_t denote the quarterly average of year t , $t = 1, 2, \dots, n$, there are

$$\bar{y}_1 = \frac{1}{4}(y_1 + y_2 + y_3 + y_4),$$

$$\dots$$

$$\bar{y}_n = \frac{1}{4}(y_{4n-3} + y_{4n-2} + y_{4n-1} + y_{4n})$$
(19)

(c) Adjust the seasonal index of each season. Theoretically, the sum of seasonal indices should be 4, but due to calculation errors in practice, the sum of seasonal indices is greater than or less than 4, so it needs to be readjusted. The adjustment formula is

$$a_i = \frac{n}{\left(\sum_{j=1}^n \bar{a}_j\right)} \bar{a}_j$$
(20)

When using the seasonal index method to forecast time series, it should be noted that the time series should not have an obvious linear trend; otherwise, the forecast accuracy will be greatly reduced.

(2) Long-term trend

This article uses the least square method to find the parameters in the linear trend formula. The core idea of the least square method is to use a straight line to approximate the historical data in the past. The mathematical language

TABLE 5: Comparison of sales forecast results and actual values of the combined forecasting model.

Month	Trending method to predict sales	Combination model predicts sales	Actual sales
201801	452738	547282	572639
201802	246371	337625	302746
201803	382736	336481	347290
201804	248474	321746	342699
201805	407634	346289	368270
201806	336585	312648	352890

is the actual observation value of the time series of the object model y_i and the predicted value \hat{y}_i in the linear trend model have the smallest sum of squared deviations, that is, the value of $\sum (y_i - \hat{y}_i)^2$ is the smallest. The least square method is used to determine the value of the parameter. The specific derivation process is omitted here. The calculation formulas for parameters a and b are

$$\hat{b} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x}.$$
(21)

(3) Decomposition prediction model

After determining the seasonal index and long-term trend with the decomposition method, the two key factors, the new forecast value of the cigarette sales model can be calculated according to formula (22).

$$X_t = T_t \times S_t \times C_t \times I_t,$$
(22)

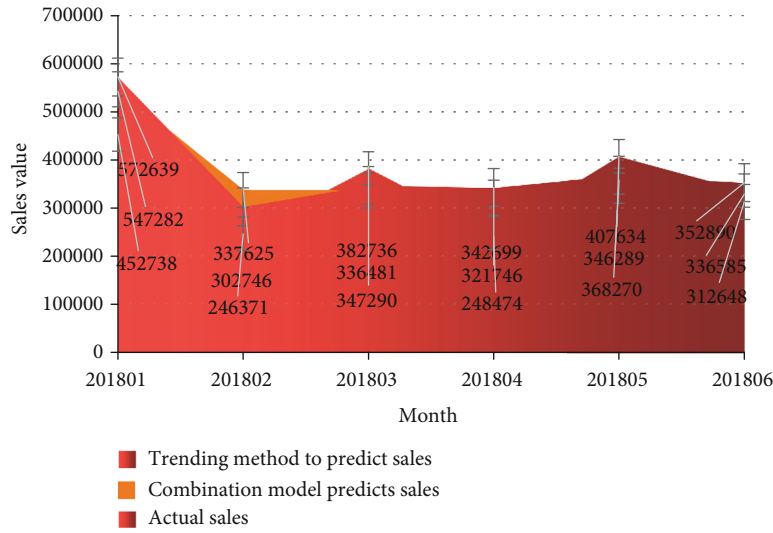


FIGURE 6: Comparison of sales forecast results and actual values of the combined forecasting model.

As the factors in the model are determined, the random fluctuation I and the cyclic index C have been reduced. After simplification, the predicted value is obtained.

4. Based on Big Data Influencing Factors of Enterprise Market Sales and Forecast Statistical Experimental Analysis

4.1. *Macroscopic Factors Affecting Cigarettes.* As shown in Table 1, among the demographic factors in this area are the resident population, floating population, and other influencing factors, cigarette sales, and other data values and data trends.

As shown in Figure 2, there are many factors affecting cigarette sales. All the factors affecting the product are analyzed by big data technology. After a large number of calculations and modeling, the product sales amount is mainly reflected in the number of permanent residents and economic efficiency is getting higher and higher over time.

4.2. *Personal Consumption Ability and Product Sales Trend.* As shown in Table 2, the data association trends between factors are per capita consumption expenditure, average wages of employees, and cigarette sales in the region.

As shown in Figure 3, according to big data, the main reason for affecting product sales is the city's per capita living expenses, followed by per capita GDP in 2016-2019. Due to the limitations of data statistics, there is a lack of regularity. In this case, traditional mathematical induction statistical methods cannot be used.

4.3. *Sales Forecast Analysis.* We use this model to obtain the sales forecast value of the key brand from January to June 2018 and compare it with the actual sales value and compare it with the sales value of the same period in 2017. The unit is box. The results are shown in Table 3.

As can be seen in Figure 4, the sales forecast model based on the trend method has a good forecast of the seasonality

TABLE 6: Comparison of prediction effects of three prediction models.

Predictive model	MAE	RMSE	MAPE (%)
Trend model	4628.3	5835.2	13.28
BP neural network model	18636	20351	7.37
Combined forecasting model	9635	113634	3.12

and periodicity of the monthly sales of cigarettes, but the relative error of the forecasted sales in a certain two months is still more than 10%. The prediction effect of extreme values in the time series is not ideal, and we can continue to improve on the basis of this prediction.

4.4. *Forecast by Seasonal Index Method.* After repeated training, the number of hidden layer nodes is determined to be 5, and the trained seasonal index method model is used to predict the test data set. The results obtained are shown in Table 4:

It can be seen intuitively from Figure 5 that the relative error of the sales forecast model based on the seasonal index method is relatively stable and the error is small. Therefore, this paper adopts a seasonal index method model based on the forecast based on the trend method model to modify the forecast value of the trend method model to improve the accuracy of the forecast.

4.5. *Establish a Combined Forecasting Model Based on Trend Method and Seasonal Index Method.* First, the linear structure part of the time series is fitted with a trend method model to obtain the predicted value \hat{L}_t . The first order and seasonal differences have been performed on the time series before modeling. The purpose is to eliminate the trend of the time series and reduce the time series. The seasonal index method is used to identify the nonlinear part e_t of the time series, and the prediction result \hat{N}_t is obtained. After repeated trials and comparisons, this paper constantly adjusts the

TABLE 7: Algorithm performance comparison.

Sample	Algorithm		Random block KNN		Classic KNN	
	Correct rate	Time	Correct rate	Time	Correct rate	Time
1	0.9315	58.23	0.8642	128.39	0.9362	382.28
2	0.9163	413.56	0.9017	983.01	0.9276	3728.28
3	0.9063	502.42	0.8437	1273.52	0.9117	5583.63

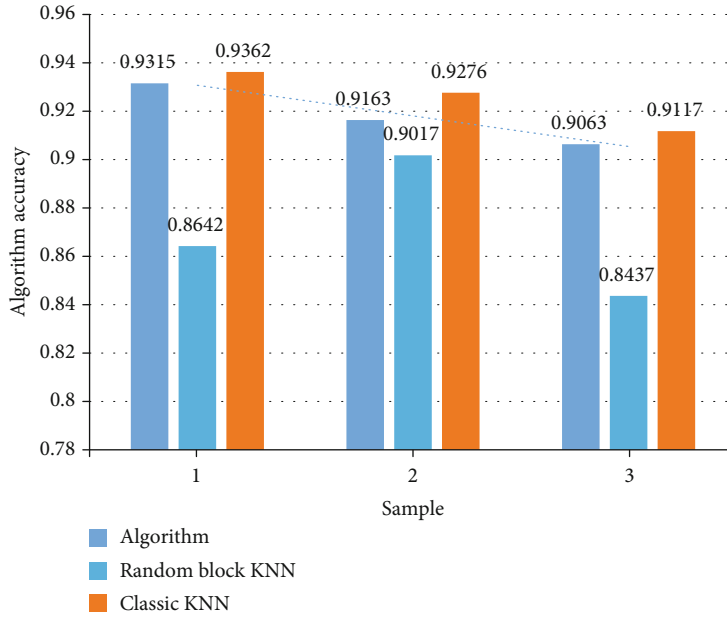


FIGURE 7: Algorithm performance comparison.

number of nodes, and finally determines the input node is 12, the hidden layer node is 5, and the output node is 1 structure. The predicted results are shown in Table 5.

It can be seen intuitively from Figure 6 that the relative error of the forecast results of the combined forecasting model is less than 5%, which is also lower than the relative error of the forecasting results of the trend forecasting model and the seasonal index forecasting model, that is, the forecasting effect is better than that of a single model. Through the comprehensive application of the two models of trend method and seasonal index method, they can give full play to their respective strengths to achieve the purpose of improving the forecasting effect.

This article uses historical monthly sales data of a certain brand of cigarettes from 2016 to 2019 to establish a trend-based sales forecast model, a seasonal index-based forecast model, and a combined forecast model to predict the monthly sales in 2018. Table 6 shows the comparison of the prediction effects of the three prediction models by the MAE, RMSE, and MAPE prediction evaluation standards.

Table 6 shows that from the three indicators of MAE, RMSE, and MAPE, it can be seen that the evaluation indicators of the combined model are the lowest among the three models, so the prediction effect is the best. In the original data, there are both linear factors and nonlinear factors, so

TABLE 8: Comparative analysis of cigarette sales volume forecast and actual results.

Time	Trend forecast	Seasonal index forecast	Decomposition prediction	Actual value
2018.6	15.8	15.37	15.31	16.24
2018.7	15.9	16.35	16.47	17.24
2018.8	17.1	15.83	16.74	16.68
2018.9	17.0	15.35	16.43	17.19
2018.10	16.8	15.43	17.53	16.83
2018.11	16.6	14.79	17.43	17.52

a single forecasting model, whether it is a trend forecasting model or a seasonal index forecasting model, cannot achieve the ideal forecasting effect. The combined model of the trend method and the seasonal index method can synthesize the advantages of a single model, better dig out the complex linear and nonlinear features behind the data, and also improve the prediction accuracy of the model.

4.6. Comparative Analysis of Algorithm Performance. Based on the above classification results, these sample data are processed into feature vector values. The KNN classifier is used in MATLAB to classify the three data samples, and the

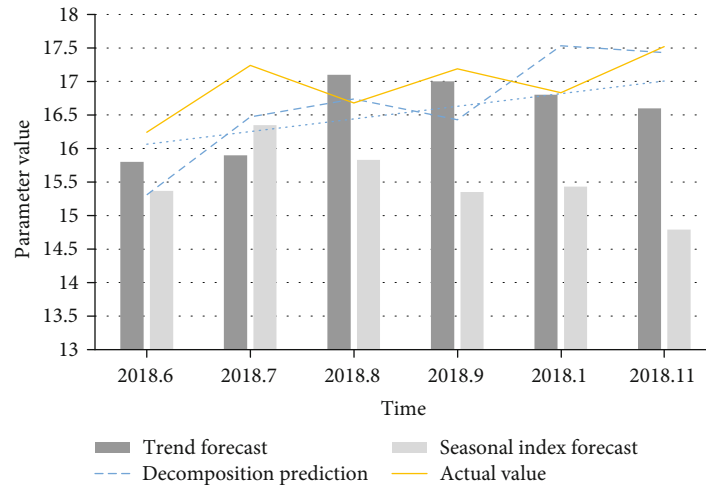


FIGURE 8: Comparative analysis of cigarette sales forecast and actual results.

number of correct classification results of the data classified by the three classification algorithms is counted. Calculate the accuracy rate obtained in each classification algorithm in each sample and use the highest accuracy rate as the accuracy rate of the classification algorithm in the sample classification result of the classification algorithm. Another performance test standard is the classification time on each sample. The results are shown in Table 7.

As shown in Table 7 and Figure 7, it can be concluded that the classification accuracy of the algorithm proposed in this paper is 0.41% ~6.2% higher than the random block algorithm, and 0.7% ~1.8% lower than the traditional KNN algorithm. It is 55.15% ~63.17% faster than the random block algorithm, and 83.28% ~90.55% faster than the traditional KNN algorithm. As the data increases, the speed will increase more obviously.

4.7. Comparative Analysis of Forecast Results. According to the data in the above chart, the value obtained by the long-term trend method predicted by the formula, and the value obtained by the decomposition method, the forecast data for the next few periods are, respectively, predicted. Finally, we compare these data with the actual sales results of the cigarette market. For comparison, the data details are shown in Table 8.

From Figure 8, we can see that the error analysis between the forecasted value of the three forecasting models of trend method, seasonal index method, and decomposition method and actual sales volume, we can see that the time series decomposition method established by the multiplication model performs best. The average error rate is about 2% and the fluctuation is small, followed by the seasonal index method, and the trend method has the worst performance of about 4% and the fluctuation is large. For the tobacco industry, the decomposition prediction model can fully meet its forecasting needs, so as to guide industrial companies to produce cigarettes and commercial companies to sell cigarettes based on the predicted values.

5. Conclusion

The KNN classification algorithm based on overlapped k -means clustering proposed in this paper is still lower than the traditional KNN algorithm in classification accuracy. This situation comes from the algorithm's clustering process, which affects the accuracy of the classification algorithm. If the effect of clustering can be improved in the future, then the accuracy of classification will catch up with the traditional classic KNN algorithm, especially if other excellent clustering algorithms can be introduced or the conditions and methods of partitioning can be changed according to the data. The effect and efficiency of the algorithm can be further improved.

Using the big data method in intelligent IoT to analyze the historical sales data of the company, its time series presents the characteristics of dual trend changes. According to its characteristics, a combined forecasting model based on the trend method and seasonal index method is proposed, and MAE, RMSE, and MAPE forecasting evaluation standards are used to compare the combined forecasting model. A comparative analysis with the forecasting effect of a single-trend forecasting model and a seasonal index model proves that the combined forecasting model is better than a single model.

This paper studies the construction of a Hadoop massive data processing platform, an in-depth study of its key technologies HDFS and MapReduce, analyzes its working mechanism, connects with the actual situation of the enterprise, analyzes the feasibility of technology implementation and environment construction, and uses the Hadoop platform to sell mass sales first data information undergoes preprocessing such as data cleaning, dimensionality reduction, and structural standardization and then provides these processed data to a relational database to perform data analysis and processing of related businesses. This provides support for sales forecasting models based on massive data processing.

Data Availability

This article selects the monthly sales volume of a certain brand of cigarettes from China Tobacco in this province from 2016 to 2019 as the analysis data set.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This paper is not supported by a special funding but by University of Queensland.

References

- [1] J. Chen, Z. Lv, and H. Song, "Design of personnel big data management system based on blockchain," *Future Generation Computer Systems*, vol. 101, pp. 1122–1129, 2019.
- [2] T. J. Chemmanur, G. Hu, and J. Huang, "Institutional investors and the information production theory of stock splits," *Journal of Financial and Quantitative Analysis*, vol. 50, no. 3, pp. 413–445, 2015.
- [3] N. Singh and S. R. Mohanty, "A review of price forecasting problem and techniques in deregulated electricity markets," *Journal of Power and Energy Engineering*, vol. 3, no. 9, pp. 1–19, 2015.
- [4] E. Yadegaridehkordi, M. Hourmand, M. Nilashi, L. Shuib, A. Ahani, and O. Ibrahim, "Influence of big data adoption on manufacturing companies' performance: an integrated DEMATEL-ANFIS approach," *Technological Forecasting and Social Change*, vol. 137, pp. 199–210, 2018.
- [5] Y. Wang and X. Han, "Research on power system load forecasting based on classification of influence factors," *Journal of Computational and Theoretical Nanoscience*, vol. 13, no. 12, pp. 9798–9803, 2016.
- [6] D. Gu, S. Khan, I. U. Khan, and S. U. Khan, "Understanding mobile tourism shopping in Pakistan: an integrating framework of innovation diffusion theory and technology acceptance model," *Mobile Information Systems*, vol. 2019, Article ID 1490617, 18 pages, 2019.
- [7] T. Xin, C. Kai, and L. I. Gang, "Influencing factors analysis and trend forecasting of China's carbon emissions—empirical study based on STIRPAT and GM (1, 1) models," *Journal of Northeastern University*, vol. 36, no. 2, pp. 297–300, 2015.
- [8] Z. Xing, H. Guo, and Q. Fu, "Analysis of influencing factors of rainfall in irrigation area and combining rainfall forecasting," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 46, no. 8, 2015.
- [9] J. Šindelář, "Investigation of factors influencing employee performance," *International Journal of Organizational Analysis*, vol. 24, no. 2, pp. 340–368, 2016.
- [10] K. Yang, "The construction of sports culture industry growth forecast model based on big data," *Personal and Ubiquitous Computing*, vol. 24, no. 1, pp. 5–17, 2020.
- [11] 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. Feb., pp. 452–462, 2020.
- [12] E. W. K. See-To and E. W. T. Ngai, "Customer reviews for demand distribution and sales nowcasting: a big data approach," *Annals of Operations Research*, vol. 270, no. 1–2, pp. 415–431, 2018.
- [13] S. Li and B. Wang, "Hybrid parallel Bayesian network structure learning from massive data using MapReduce," *Journal of Signal Processing Systems*, vol. 90, no. 8–9, pp. 1115–1121, 2018.
- [14] P. G. Kulkarni and S. R. Khonde, "An improved technique of extracting frequent itemsets from massive data using MapReduce," *International Journal of Engineering and Technology*, vol. 9, no. 3S, pp. 400–406, 2017.
- [15] S. S. Gavankar and S. D. Sawarkar, "Eager decision tree," in *2017 2nd International Conference for Convergence in Technology (I2CT)*, pp. 837–840, Mumbai, India, 2017.
- [16] D. Wang, D. Yuan, and C. Miao, "Sparse naïve Bayes base on entropy correlation for GPR image denoising," in *2020 IEEE 3rd International Conference on Electronics and Communication Engineering (ICECE)*, pp. 167–171, Xi'an, China, 2020.
- [17] Z. Liu and L. Bai, "Evaluating the supplier cooperative design ability using a novel support vector machine algorithm," in *2008 12th International Conference on Computer Supported Cooperative Work in Design*, pp. 986–989, Xi'an, China, 2008.
- [18] J. Vieira, R. P. Duarte, and H. C. Neto, "kNN-STUFF: kNN STreaming unit for Fpgas," *IEEE Access*, vol. 7, pp. 170864–170877, 2019.
- [19] J. H. Andreae, "Brains, neural networks and expert systems," in *Proceedings 1993 The First New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems*, pp. 3–4, Dunedin, New Zealand, 1993.
- [20] Y. Ma, X. Meng, and S. Wang, "Parallel similarity joins on massive high-dimensional data using MapReduce," *Concurrency and Computation: Practice and Experience*, vol. 28, no. 1, pp. 166–183, 2016.
- [21] A. T. Azar and A. E. Hassaniien, "Dimensionality reduction of medical big data using neural-fuzzy classifier," *Soft Computing*, vol. 19, no. 4, pp. 1115–1127, 2015.
- [22] F. N. Afrati, S. Sharma, J. R. Ullman, and J. D. Ullman, "Computing marginals using MapReduce," *Journal of Computer and System Sciences*, vol. 94, pp. 98–117, 2015.
- [23] S. K. Zhang, G. Y. Shi, Z. J. Liu, Z. W. Zhao, and Z. L. Wu, "Data-driven based automatic maritime routing from massive AIS trajectories in the face of disparity," *Ocean Engineering*, vol. 155, pp. 240–250, 2018.
- [24] H. Duan, Y. Peng, G. Min, X. Xiang, W. Zhan, and H. Zou, "Distributed in-memory vocabulary tree for real-time retrieval of big data images," *Ad Hoc Networks*, vol. 35, pp. 137–148, 2015.
- [25] D. Wang and J. Liu, "Optimizing big data processing performance in the public cloud: opportunities and approaches," *IEEE Network*, vol. 29, no. 5, pp. 31–35, 2015.
- [26] C. Sreedhar, N. Kasiviswanath, and P. Chenna, "A survey on big data management and job scheduling," *International Journal of Computer Applications*, vol. 130, no. 13, pp. 41–49, 2015.
- [27] 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. 2, pp. 186–209, 2015.
- [28] Byung, Ho, Jung, Dong, Hoon, and Lim, "RHadoop," *Journal of the Korea Society of Computer and Information*, vol. 22, no. 4, pp. 9–16, 2017.

- [29] A. Pandey, "Simplilearn big data hadoop review," *PC Quest*, vol. 32, no. 4, pp. 33–33, 2019.
- [30] M. Grossman, M. Breternitz, and V. Sarkar, "HadoopCL2: motivating the design of a distributed, heterogeneous programming system with machine-learning applications," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 3, pp. 762–775, 2016.
- [31] K. McDermott, "Achieving data liquidity across health care requires a technical architecture," *Bulletin of the Association for Information Science and Technology*, vol. 43, no. 1, pp. 19–22, 2016.
- [32] Z. Lv, D. Chen, and A. K. Singh, "Big data processing on volunteer computing," *ACM Transactions on Internet Technology (TOIT)*, 2020.

Research Article

Indoor PDR Positioning Assisted by Acoustic Source Localization, and Pedestrian Movement Behavior Recognition, Using a Dual-Microphone Smartphone

Mei Wang,^{1,2} Nan Duan,¹ Zou Zhou ,^{1,3} Fei Zheng ,^{1,3} Hongbing Qiu ,^{1,3} Xiaopeng Li,¹ and Guoli Zhang¹

¹School of Information and Communication, Guilin University of Electronic Technology, Guilin 541004, China

²College of Information Science and Engineering, Guilin University of Technology, Guilin 541004, China

³Ministry of Education Key Laboratory of Cognitive Radio and Information Processing, Guilin University of Electronic Technology, Guilin 541004, China

Correspondence should be addressed to Zou Zhou; zhouzou@guet.edu.cn and Fei Zheng; zhengfei@guet.edu.cn

Received 26 March 2021; Accepted 15 June 2021; Published 9 July 2021

Academic Editor: Peng Yu

Copyright © 2021 Mei 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 recent years, the public's demand for location services has increased significantly. As outdoor positioning has matured, indoor positioning has become a focus area for researchers. Various indoor positioning methods have emerged. Pedestrian dead reckoning (PDR) has become a research hotspot since it does not require a positioning infrastructure. An integral equation is used in PDR positioning; thus, errors accumulate during long-term operation. To eliminate the accumulated errors in PDR localisation, this paper proposes a PDR localisation system applied to complex scenarios with multiple buildings and large areas. The system is based on the pedestrian movement behavior recognition algorithm proposed in this paper, which recognises the behavior of pedestrians for each gait and improves the stride length estimation for PDR localisation based on the recognition results to reduce the accumulation of errors in the PDR localisation algorithm itself. At the same time, the system uses self-researched hardware to modify the audio equipment used for broadcasting within the indoor environment, to locate the acoustic source through a Hamming distance-based localisation algorithm, and to correct the estimated acoustic source estimated location based on the known source location in order to eliminate the accumulated error in PDR localisation. Through analysis and experimental verification, the recognition accuracy of pedestrian movement behavior recognition proposed in this paper reaches 95% and the acoustic source localisation accuracy of 0.32 m during movement, thus, producing an excellent effect on eliminating the cumulative error of PDR localisation.

1. Introduction

In recent years, the public's demand for location-based services (LBS) has become more robust, and LBS has affected many aspects of people's work and life. The Global Navigation Satellite System (GNSS) is a significant component of LBS, and the accuracy has reached the submeter level [1, 2]. Although GNSS can provide good coverage and high accuracy outdoors, it does not meet the positioning requirements indoors and in other sheltered environments due to the low satellite signal strength and quality in these locations [3]. Therefore, there is an urgent need for indoor LBS with real-

time, stable, and accurate positioning performance, especially in schools, hospitals, and large commercial shopping centers due to the large and complex area and numerous users [4].

Commonly used indoor positioning technology can be divided into three categories based on the positioning principle: trilateral positioning [5], fingerprint positioning [6], and track estimation [7]. In complex scenarios, these three indoor positioning technologies can provide relatively good positioning performance. However, trilateral positioning requires a receiver and a reference source, and it is necessary to deploy the equipment in the area in advance. Therefore, full coverage is required in the facility to prevent blind spots,

TABLE 1: Comparison of the latest indoor positioning methods.

Solution	Principle	Precision (m)	Dimension	Coverage area (m ²)	Cost (\$/m ²)	Compatible with smartphone
Bluetooth	Trilateral positioning	<2	2D	30	0.4	Yes
WiFi	Fingerprint positioning	<2	2D	50	0.6	Yes
PDR	Track estimation	<4	2D	/	/	Yes
UWB	Trilateral positioning	<0.04	3D	100	60	No
Acoustics	Trilateral positioning	<0.3	3D	60	0.9	Yes

increasing the positioning cost. Fingerprint positioning can achieve good positioning accuracy in scenarios without or sparsely deployed positioning facilities. In recent years, WiFi-based fingerprint positioning has attracted attention due to its deployment practicability. However, the construction and update of fingerprint maps is a complex and tedious task. Besides, a fingerprint map has the problem of timeliness [6], which reduces the positioning accuracy and substantially increases the positioning cost. The typical trajectory calculation principle in indoor positioning is pedestrian dead reckoning (PDR) positioning. Levi and Judd proposed PDR positioning in 1996 [7]; it has been widely used in complex environments because of its advantages, such as no infrastructure. It achieves good positioning accuracy.

We compared the latest indoor positioning solutions reported in the literature [8–12] regarding positioning accuracy, coverage, cost, smartphone compatibility, and other parameters, as shown in Table 1.

The comparison shows the advantages of PDR over other indoor positioning solutions, but it is rarely used alone for positioning in complex scenes. The main reason is that PDR positioning has good short-distance positioning accuracy, but errors are accumulated due to long-term operation. Most scholars have researched these two aspects to reduce the cumulative error in PDR positioning and improve the positioning accuracy. One solution is to combine PDR with other positioning methods to improve the positioning accuracy, such as Bluetooth low energy (BLE)/PDR fusion [13], PDR/WiFi fusion [14], PDR/WiFi/geomagnetic fusion, and other methods [15]. However, positioning infrastructure is required, which increases and is not conducive to large-scale applications. Another solution is to reduce the cumulative error using step detection, stride length estimation, and heading determination to reduce the impact of noise [16]. This approach reduces the error accumulation but fails to eliminate the root cause of the error. In addition, in previous PDR positioning research, the default smartphone was rigidly linked to the user, i.e., the user holds the smartphone for PDR positioning. Numerous studies were conducted on the smartphone's position and attitude and user movement in PDR positioning [17, 18]. Although many optimization methods exist for PDR positioning, we believe that the optimal PDR positioning method should reduce the error accumulation and the cost to achieve the most practical and cost-effective solution.

In addition to optimizing the PDR positioning method, it is also necessary to investigate the infrastructure necessary for indoor positioning applications, such as schools, hospi-

tals, and large commercial shopping centers. Besides the standard public WiFi and visible light equipment, audio playback equipment used for broadcasting is often ignored. Audio equipment is widely deployed in indoor locations, supporting the positioning method proposed in this article. This article proposes a PDR positioning method based on an acoustic source for positioning correction using dual-microphone smartphones. The main contributions of this article are as follows:

- (1) We propose a practical and cost-effective indoor positioning method that is suitable for multiple buildings and large areas. The method introduces pedestrian movement behavior recognition to improve PDR localisation accuracy, while using acoustic source localisation to reverse the cumulative error of PDR localisation
- (2) We propose an acoustic localisation algorithm based on Hamming distance. The algorithm uses pseudoultrasonic sound, which is inaudible to the human ear and can be processed by ordinary speakers and microphones, at frequencies between 18 kHz and 20 kHz as the localisation signal. And use the known position of the sound source to correct the estimated position, in order to achieve the purpose of eliminating the accumulated error of PDR positioning
- (3) We propose a method to improve the accuracy of PDR localisation using pedestrian movement behavior recognition. The method uses the proposed method based on gait periodicity features proposed in this paper to extract features from the data collected by smartphones, uses support vector machine (SVM) as a classifier to recognize the movement behavior of each gait, and uses Dempster–Shafer (D-S) evidence theory to fuse the recognition results for the problem of low recognition accuracy in certain complex scenes, improving the overall recognition accuracy to 96%. Based on the excellent recognition accuracy, this paper improves the stride length estimation for PDR localization and eliminates the cumulative error of PDR localization using map matching

This paper is organized as follows: Section 2 describes the related work of PDR positioning, pseudoultrasonic positioning, and acoustic source positioning based on dual-microphone smartphones. Section 3 presents the proposed

system framework. In Section 4, we introduce the methods, and Section 5 describes the experiments. Section 6 provides the conclusion and future work.

2. Related Works

PDR positioning requires the establishment of an integral equation; thus, minimizing or eliminating the accumulated errors has been a research focus. Studies have shown [13–15] that fusion methods have several advantages and can significantly reduce the accumulated error of PDR positioning. The disadvantage is that the infrastructure needs to be deployed in advance, which requires access to the facilities and increases the costs. The optimization of the PDR algorithm includes the optimization of the step detection part [19, 20] or the continuous iteration of the step estimation model [15, 21, 22]. The advantage of PDR optimization is the minimization of the error generated in each step, which reduces the cumulative error. However, the disadvantage is that the cumulative error cannot be entirely eliminated.

Scholars have used PDR positioning in increasingly complex scenarios in addition to efforts to eliminate the accumulated error. Most previous PDR studies assumed that the smartphone remained in the same position (maintain a horizontal hand-held position) at different pedestrian locations, the person and the smartphone were rigidly connected, and the positioning area was two-dimensional. In real-world conditions, the behavior of pedestrians is complex and variable. In other words, pedestrians may walk, run, or go up and downstairs in an indoor environment, and the smartphone may be placed horizontally in front of the chest, held by the hand, or placed in a pocket. Researchers often ignore these variable conditions, resulting in a single PDR positioning scenario and significant use limitations. Dirican and Aksoy [23] proposed a spectrum analysis method based on fast Fourier transform (FFT) for step detection to convert the time-domain information of the step frequency to the frequency domain to deal with more complex application scenarios. However, this approach requires a fixed time window, making it impossible to identify walking at different speeds accurately. Mohssen et al. [24] analyzed scenarios where the pedestrian heading direction was inconsistent with the device and proposed a heading estimation model based on principal component analysis. This method improved the PDR positioning accuracy in different scenarios. However, only a two-dimensional structure was considered.

Scholars have also used machine learning methods to solve this problem due to recent developments in this field. References [11, 17, 25, 26] used different classification methods, such as K -nearest neighbor, dynamic time warping, SVMs, neural networks, decision trees, and hidden Markov model to analyze acceleration, angular velocity, and magnetic field strength related to pedestrian behavior recognition. Wang et al. [17] separated the recognition of pedestrian movement/behavior into two steps: recognition of the smartphone pose (hold, calling, swinging, and pocket) and recognition of the pedestrian movement (walk, run, upstairs, and downstairs). SVM and decision tree were used to detect the

pedestrian movements and smartphone pose with a recognition rate of 92.4%. The authors used only PDR positioning and reduced the positioning error in a 164 m trajectory to 3.5 m for different pedestrian movements and smartphone poses.

The improvement in the PDR positioning accuracy for pedestrian movement recognition requires optimization of the PDR algorithm. Therefore, it is essential to develop low-cost positioning methods and eliminate the accumulated error of PDR. Acoustic positioning has a lower equipment cost than other smartphone-based positioning methods. Therefore, this approach has attracted the interest of many research teams. Acoustic-based indoor positioning can be divided into three categories: time of arrival (TOA), time difference of arrival (TDOA), and difference of arrival (DOA) methods. In the past ten years, several acoustic positioning methods have emerged, such as the ASSIST positioning system proposed by Höflinger et al. in 2012 [27]. It uses acoustic signals in the 18 kHz–21 kHz frequency band and the TDOA positioning principle to increase the positioning accuracy to 0.26 m. The GuoGuo positioning system proposed by Liu et al. [28] in 2015 uses acoustic signals in the 17 kHz–22 kHz frequency band and the TOA positioning principle to increase the positioning accuracy to 0.25 m in a static scene. In the past two years, the research group led by Wang et al. at Zhejiang University used acoustic signals in the frequency band of 20 kHz, resulting in high accuracy [12, 29]. Acoustic signals in the frequency band of 20 kHz have unique ultrasonic characteristics, i.e., strong anti-interference ability, as well as broader transmission than other ultrasonic wavelengths. Since ultrasonic wavelengths are not perceived by humans, there is no danger of noise pollution. Therefore, pseudoultrasound is suitable for use as positioning signals.

Few studies were conducted on acoustic source localization using dual-microphone smartphones. Acoustic source localization is different from acoustic-based indoor positioning in that acoustic source localization is used to locate the location of the acoustic emitting source, while acoustic-based indoor positioning uses acoustic as a positioning signal for information transmission to achieve the purpose of localization. These two are essentially different elements. The acoustic source localization principle is to use the dual microphones of smartphones to create a microphone array. The localization method is the same as for acoustic source localization of a microphone array. For example, Jin et al. [30, 31] used a dual-microphone smartphone, a Hamming window, and weighted probability to achieve acoustic source localization accuracy of 0.19 m in a static scene. This study uses a dual-microphone smartphone for acoustic source localization in dynamic scenarios to eliminate the cumulative error of PDR localization.

3. System Overview

The proposed system framework consists of three parts: pedestrian movement behavior recognition, PDR positioning, and PDR positioning correction. The smartphone represents the system input, and the current pedestrian position is the output. The system block diagram is shown in Figure 1.

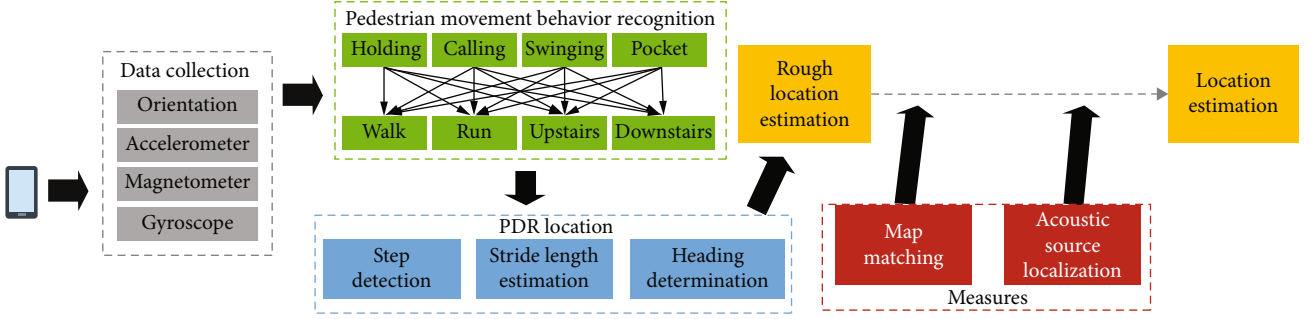


FIGURE 1: The system block diagram. The data are collected by the smartphone, and the estimated position is the output.

In PDR positioning, the estimated position of the k -th step of the walking pedestrian can be expressed as an integral equation:

$$\begin{bmatrix} \hat{x}_k \\ \hat{y}_k \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + \sum_{i=1}^k \hat{L}_i \cdot \begin{bmatrix} \sin(\hat{\theta}_i) \\ \cos(\hat{\theta}_i) \end{bmatrix}, \quad (1)$$

where $\begin{bmatrix} \hat{x}_k \\ \hat{y}_k \end{bmatrix}$ represents the estimated position of the k -th step, $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$ represents the initial position, \hat{L}_i is the estimated step length of the i -th step, and $\hat{\theta}_i$ is the estimated heading angle of the i -th step. The cumulative error is expressed independently by rewriting Equation (1) as:

$$\begin{bmatrix} \hat{x}_k \\ \hat{y}_k \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + \sum_{i=1}^k L_i \cdot \begin{bmatrix} \sin(\theta_i) \\ \cos(\theta_i) \end{bmatrix} + \sum_{i=1}^k \begin{bmatrix} \Delta x_i \\ \Delta y_i \end{bmatrix}, \quad (2)$$

where L_i represents the step length of the i -th step, and $\begin{bmatrix} \Delta x_i \\ \Delta y_i \end{bmatrix}$ represents the position estimation error of the i -th step. In this article, we consider the primary error source of the position estimation the step size estimation term when the heading angle estimation error is small; thus,

$$\min \sum_{i=1}^k \begin{bmatrix} \Delta x_i \\ \Delta y_i \end{bmatrix} = \min \sum_{i=1}^k \Delta L_i \cdot \begin{bmatrix} \sin(\theta_i) \\ \cos(\theta_i) \end{bmatrix}. \quad (3)$$

The proposed pedestrian movement behavior recognition method adaptively estimates the step length for 16 combinations of pedestrian movements and smartphone poses, reducing the step length term's cumulative error.

Although the recognition of the pedestrian movement behavior can reduce the cumulative error of PDR positioning, the error is not entirely eliminated. Therefore, we incorporate two global optimization items: map matching and acoustic source localization. Therefore, Equation (1) can be expressed as:

$$\begin{bmatrix} \hat{x}_k \\ \hat{y}_k \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + \sum_{i=1}^k L_i \cdot \begin{bmatrix} \sin(\theta_i) \\ \cos(\theta_i) \end{bmatrix} + \sum_{i=1}^k \Delta L_i \cdot \begin{bmatrix} \sin(\theta_i) \\ \cos(\theta_i) \end{bmatrix} - \sum_{i=1}^n \begin{bmatrix} x_L \\ y_L \end{bmatrix} - \sum_{i=1}^p \begin{bmatrix} x_A \\ y_A \end{bmatrix}, \quad (4)$$

where $\sum_{i=1}^n \begin{bmatrix} x_L \\ y_L \end{bmatrix}$ represents the position coordinate parameter that needs to be corrected after n times of map matching, $\sum_{i=1}^p \begin{bmatrix} x_A \\ y_A \end{bmatrix}$ represents the position coordinate parameter that needs to be adjusted after p times of acoustic source positioning correction.

4. Pedestrian Dead Reckoning (PDR)

The PDR system comprises three parts: step detection, stride length estimation, and heading determination, and all three parts contain errors. Most previous studies assumed that the smartphone was rigidly connected to the human body. However, this state does not apply in actual scenarios. This section describes the method of using pedestrian movement behavior recognition to improve the stride length estimation accuracy of PDR positioning. In addition, a map matching algorithm is used to distinguish moving upstairs and downstairs to eliminate the cumulative error of the PDR.

5. Movement Behavior Recognition

The proposed pedestrian movement behavior recognition method is described in detail in one of our recent conference papers. This section describes the details of this method. We divide the pedestrian movement behavior recognition into smartphone pose recognition and pedestrian movement recognition. We believe that a nonmoving pedestrian does not have to be considered in indoor positioning because the sensor of the smartphone does not record movement, and no positioning errors are created. Thus, this state will be discussed separately. The categories of pedestrian movement behavior recognition considered in this article are shown in Figure 2.

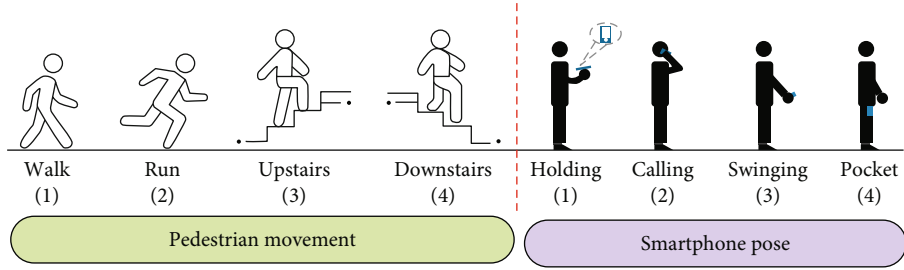


FIGURE 2: Schematic diagram of the recognition category of the pedestrian movement behavior. Four pedestrian movements and four smartphone poses are considered in pairs, resulting in 16 pedestrian movement behavior types.

5.1. Smartphone Pose Recognition. Compared with PCs, smartphones have limited computing capabilities. Therefore, complex machine learning algorithms used in smartphone pose recognition will significantly increase the system's calculations, which is not conducive to the fast and stable operation of the system. The internal resources of smartphones are often not considered for implementing functions. For example, Android smartphones have many built-in sensors, referred to as hardware sensors. Therefore, the Android kernel functions are accessible; those are referred to as software sensors. A software sensor performs calculations using data obtained from hardware sensors. An example is the angle sensor.

On the Android platform, the angle sensor is called "TYPE_ORIENTATION." The official document shows that it provides the device's pitch and rotation (roll). Therefore, we can use clustering to obtain the smartphone pose, as expressed in Equation (5):

$$\text{Phone Pose} = \begin{cases} \text{Holding, } x_1^1 < \text{pitch} < x_1^2, y_1^1 < \text{pitch} < y_1^2; \\ \text{Calling, } x_2^1 < \text{pitch} < x_2^2, y_2^1 < \text{pitch} < y_2^2; \\ \text{Swinging, } x_3^1 < \text{pitch} < x_3^2, y_3^1 < \text{pitch} < y_3^2; \\ \text{Pocket, } x_4^1 < \text{pitch} < x_4^2, y_4^1 < \text{pitch} < y_4^2, \end{cases} \quad (5)$$

where x_i^1 and x_i^2 ($i = 1, 2, 3, 4$) represent the upper and lower limits of the pitch angle, and y_i^1 and y_i^2 ($i = 1, 2, 3, 4$) represent the upper and lower limits of the roll angle, respectively.

5.2. Feature Extraction. We use the feature extraction method described in our previous paper [32]. It is based on the periodicity of pedestrian movement, and the acceleration and angular velocity data are segmented and transformed into feature vectors. This section discusses the use of this method in this study. The proposed method's advantages over commonly used sliding window methods for feature extraction are as follows: (a) the sliding window method commonly uses a redundant design to avoid data leakage. The proposed method does not use this approach, improving the system's operating efficiency. (b) The eigenvectors obtained after segmentation include the acceleration and angular velocity information of each gait, providing information on the movement behavior at each gait. The vector length maps the gait frequency, resulting in a high recognition rate of

the movement behavior with few dimensional features. (c) The pedestrian movement behavior in each gait can be determined accurately.

5.3. State Recognition Algorithm of the Movement Transition Zone Based on SVM and D-S Evidence Theory. In complex positioning scenarios such as connected buildings, stairs are located at the buildings' junctions and between floors. Therefore, pedestrians need to change their movement states continually. The previous discussion shows that we can accurately recognize the pedestrian's current movement state. However, when the pedestrian transitions between different movements, a transition zone occurs, leading to recognition errors. We use an SVM and D-S evidence theory to reduce these errors.

Few scholars investigated the errors occurring in the transition between different states, but this error exists and reduces the positioning accuracy. Figure 3(a) shows typical stairs in connected buildings, Figure 3(b) is a schematic diagram of the longitudinal section of the stairs, and Figure 3(c) shows the acceleration curve as a pedestrian is walking on the stairs. When the pedestrians are walking on level ground and the stairs, the acceleration curve shows periodicity, and when a transition occurs between walking on level ground and the stairs, the acceleration value curve shows fewer fluctuations. Therefore, it is difficult to identify the current state of motion accurately, and we confirmed this in experiments.

Although this process is concise, and there are only two or three steps that cause errors, we believe that the stride lengths are different when the pedestrian walks downstairs or upstairs. If we mistakenly use the wrong stride length for walking downstairs, the error of each step is close to 0.5 m, significantly increasing the positioning error.

The first condition for using the map matching algorithm to eliminate the cumulative error (this is discussed in the next section) is to obtain an accurate position of the current landmark. Therefore, it is crucial in this study to detect the transition between different movement states of the pedestrian. The DS evidence theory is an uncertainty reasoning method. It uses multiple inaccurate judgments and descriptions of problems, focuses on consistent information, and eliminates and integrates contradictory information to determine inaccurate conclusions. Therefore, the combination of D-S evidence theory and SVM is well suited to identify the transition zone between different movement states. The algorithm framework of this approach has been described in our

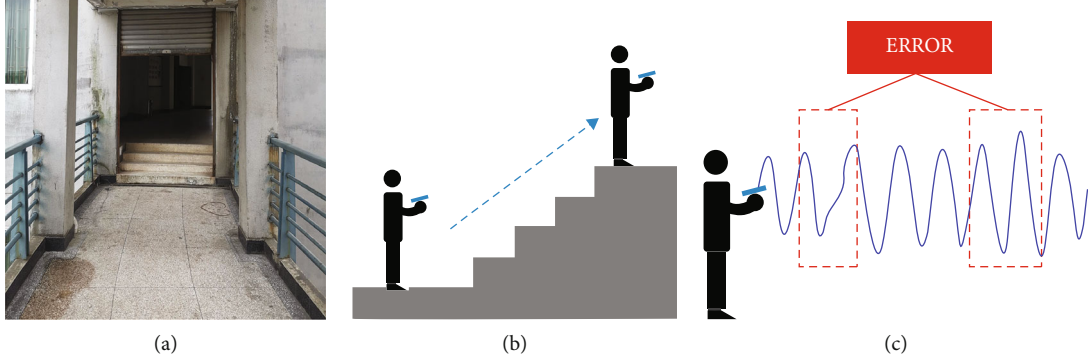


FIGURE 3: Schematic diagram of the scene. (a) Typical stairs found in connected buildings. (b) Longitudinal section view of stairs. (c) Curve of acceleration values as a pedestrian is walking on the stairs.

```

Input: The movement state of the pedestrian's current gait;
Output: Update the movement state of the first 5 steps of the pedestrian's current gait.
1: Determine the current movement state
if movement state = upstairs or downstairs
  1.1: for i =1 to 5 do:
  1.2: Fusion of all categories:
    1.2.1: Set A=Walk, B=Upstairs, C=Downstairs;
    1.2.2: for n =1 to 3 do:
      1.2.2.1:if n =1, X = A;
    else if n =2, X = B;
      else if n =3, X = C;
    end if
    1.2.2.2: Obtain  $P_j(x_i | x \in X), j = \text{SVM1 or SVM2}$ ;
    1.2.2.3: Obtain  $P_j(x_i | x \notin X), j = \text{SVM1 or SVM2}$ ;
    1.2.3.4: Obtain  $P_j(x_i | x \notin \text{Unkown}), j = \text{SVM1 or SVM2}$ ;
    1.2.3: end
  1.3: Fusion between categories:
    1.3.1: Obtain the probability  $P_i(x | x \in A), P_i(x | x \in B), P_i(x | x \in C)$ 
  1.4: Decision output.
  1.5: end
end if
2: Update the first 5 steps of the movement state

```

ALGORITHM 1: Algorithm for detecting the movement transition zone based on D-S evidence theory and SVM.

previous papers. Here, we only discuss the specific application of the algorithm. The pseudocode of the algorithm is shown in Algorithm 1.

In Algorithm 1, SVM1 in 1.2.2.2 represents a classifier constructed using acceleration data, and SVM2 represents a classifier constructed using angular velocity; $P_j(x_i | x \in X)$ represents the probability of the current event x input event X ($X = \text{Walk} \setminus \text{Upstairs} \setminus \text{Downstairs}$), j represents the classifiers difference, and n represents the events difference. $P_i(x | x \in A)$ represents the probability that the i -th step of the current output belongs to walk\upstairs\downstairs.

5.4. Stride Length Estimation Based on Pedestrian Movement Behavior Recognition. In Section 3, we listed the primary sources of position estimation errors. Equation (3) indicates that the stride length estimation error has a more significant impact on the overall error when the heading angle estimation error is small.

If the pedestrian's movement state from step j to l in k steps is the upstairs movement state when the movement state of the pedestrian is not recognized, we will mistakenly classify it as the walking state. The stride length of normal adults is 0.6 m-0.8 m during walking, whereas the stride length on the stairs is only about 0.2 m; the resulting error is usually tens of meters. Therefore, the proposed pedestrian movement/behavior recognition method has high accuracy.

The stride length estimation algorithm is continuously updated with the development of PDR technology. The most commonly used algorithm is the stride length optimal parameter estimation algorithm proposed by Shin et al. in 2011. The stride length is defined as:

$$\text{Stride Length} = \alpha \cdot f + \beta \cdot \sigma^2 + \gamma, \quad (6)$$

where f is the step frequency, i.e., $f = 1/\Delta T$; σ^2 is the acceleration mode variance; α and β are the parameter weights, and γ

```

Input: The movement state of the pedestrian's current gait.
Output: The stride length of the pedestrian's current gait.
1: Determine the  $\alpha$  and  $\beta$  parameters in different movement behavior states
switch smartphone pose:
case holding:
switch movement state:
case walk:
 $\alpha = \alpha_{hw}; \beta = \beta_{hw};$ 
case run :
 $\alpha = \alpha_{hr}; \beta = \beta_{hr};$ 
case upstairs :
 $\alpha = \alpha_{hu}; \beta = \beta_{hu};$ 
case downstairs :
 $\alpha = \alpha_{hd}; \beta = \beta_{hd};$ 
case downstairs:
 $\alpha = \alpha_{hd}; \beta = \beta_{hd};$ 
case calling:
switch movement state:
Same process 1;
case swinging:
switch movement state:
Same process 1;
case pocket:
switch movement state:
Same process 1;
end
2: Substitute  $\alpha$  and  $\beta$  into Equation (6)

```

ALGORITHM 2: Adaptive stride length estimation.

is a compensation constant. Although this algorithm is more flexible than linear and nonlinear stride length estimation algorithms, the α and β parameters are often fixed, which is not suitable for pedestrians performing complex activities.

In the pedestrian movement behavior recognition, we consider four movements (walk, run, upstairs, and downstairs) and four smartphone poses (holding, calling, swinging, and pocket). The 16 pedestrian behaviors had different parameters. The pseudocode of the stride length estimation algorithm is shown in Algorithm 2.

5.5. Cumulative Error Elimination in PDR

5.5.1. Cumulative Errors Based on Map Matching. In indoor positioning, a geometric algorithm is commonly used for

map matching. It uses the map's geometric information to extract points on the path corresponding to the position information. The geometric relationship between the current position information and the path network is analyzed [33]. There are three methods to implement geometric algorithms: point-to-point matching, point-to-arc matching, and arc-to-arc matching. Since we use line segments, point-to-arc matching is used for map matching. In this method, we find the line segment closest to the estimated point on the map, obtain the vertical line from the estimated point to the line segment, and use the vertical foot as a new estimated point. The matching process can be expressed as:

$$\text{Distance} = \frac{[X_{\text{current}}(X_2 - X_1) + Y_{\text{current}}(Y_2 - Y_1)] + (Y_2 - Y_1)(X_1 Y_2 - X_2 Y_1)}{\sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}}, \quad (7)$$

$$X_{\text{revise}} = \frac{[X_{\text{current}}(X_2 - X_1) + Y_{\text{current}}(Y_2 - Y_1)] + (Y_2 - Y_1)(X_1 Y_2 - X_2 Y_1)}{[(X_2 - X_1)^2 + (Y_2 - Y_1)^2](X_2 - X_1)^{-1}}, \quad (8)$$

$$Y_{\text{revise}} = \frac{[X_{\text{current}}(X_2 - X_1) + Y_{\text{current}}(Y_2 - Y_1)] + (Y_2 - Y_1)(X_1 Y_2 - X_2 Y_1)}{[(X_2 - X_1)^2 + (Y_2 - Y_1)^2](Y_2 - Y_1)^{-1}}, \quad (9)$$

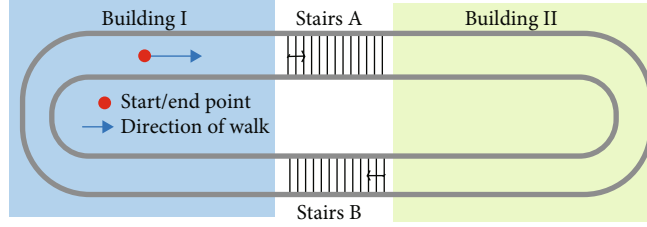


FIGURE 4: Schematic diagram of the scene. Start from the starting point and walk along the running track, passing through two stairs, to the end.

where $(X_{\text{current}}, Y_{\text{current}})$ represents the estimate of the current PDR position, and (X_1, Y_1) and (X_2, Y_2) represent the starting coordinates of a given line segment in the map database. The distance in Equation (7) is used to calculate the distance between $(X_{\text{current}}, Y_{\text{current}})$ and each line segment. The line segment with the shortest distance is obtained by projecting $(X_{\text{current}}, Y_{\text{current}})$ into the line segment using the foot position $(X_{\text{revise}}, Y_{\text{revise}})$ in Equations (8) and (9). Subsequently, the position coordinates are updated.

We assume the scene illustrated in Figure 4. Building I and Building II make up the height difference between them by means of stairs A and B. In this scenario, a pedestrian takes N steps to walk a circle and encounters stairs A and B (in m) at the k -th and p -th steps.

First, the cumulative error generated in N steps can be obtained by Equation (3):

$$\text{Error}(N) = \sum_{i=1}^N \Delta L_i \cdot \begin{bmatrix} \sin(\theta_i) \\ \cos(\theta_i) \end{bmatrix}. \quad (10)$$

For the convenience of presentation, we normalize the cumulative error generated in N steps to obtain:

$$\overline{\text{Error}} = \frac{\sum_{i=1}^N \Delta L_i \cdot \begin{bmatrix} \sin(\theta_i) \\ \cos(\theta_i) \end{bmatrix}}{N}. \quad (11)$$

After using map matching, the cumulative error within $n(0 \leq n \leq N)$ steps can be expressed as:

$$\text{Error} = \begin{cases} \overline{\text{Error}} \cdot k, & (0 \leq n \leq k); \\ \overline{\text{Error}} \cdot (p - k) + \text{Error}_{\text{map1}}, & (k \leq n \leq p); \\ \overline{\text{Error}} \cdot (N - p) + \text{Error}_{\text{map2}}, & (p \leq n \leq N), \end{cases} \quad (12)$$

where $\text{Error}_{\text{map1}}$ and $\text{Error}_{\text{map2}}$ represent the positioning error generated when the map is matched in the k -th and p -th steps. After map matching, the global cumulative error becomes the error of the endpoint:

$$\text{Error}(N) = \overline{\text{Error}} \cdot (N - p) + \text{Error}_{\text{map2}}. \quad (13)$$

Therefore, the objective of map matching is to eliminate the cumulative error at the matching point; thus, the error

depends on the number and layout of the matching points. However, the amount of landmark information in the positioning environment is limited, and the layout of landmark information is fixed. Therefore, a more flexible cumulative error elimination method is needed. In this study, we use acoustic source localization eliminate.

5.5.2. PDR with Acoustic Source Localization Correction. This correction method uses a dual-microphone smartphone as a microphone array to receive acoustic signals. Multiple signals are received by the pedestrians' smartphones, and an acoustic source localization algorithm based on the Hamming distance is used. If the sound source positioning's accuracy meets the requirements and the acoustic source's position is known, the PDR positioning can be corrected to eliminate the accumulated error.

5.5.3. Acoustic Source Data Association. There are multiple different acoustic sources in the localization environment, and when a mobile phone collects audio data, it collects data from multiple acoustic sources. Moreover, each acoustic source used to correct the PDR data has a fixed position; thus, the mobile phone has to distinguish different acoustic sources and associate them with the position information. Our solution has the following approach:

- (1) We used pseudoultrasound as the acoustic signal to increase the system's anti-interference ability. Pseudoultrasound refers to an acoustic frequency range of 18 kHz-20 kHz. Humans cannot hear this sound, but ordinary speakers and microphones can acquire the acoustic signal. The pseudoultrasonic signal has the advantages of strong anti-interference, strong privacy, and no noise interference audible to humans. Unlike ultrasonic signals, pseudoultrasonic signals can be received using standard audio equipment, no unique ultrasonic signal is needed, and the propagation distance is longer than for ultrasonic signals
- (2) We used self-developed hardware to broadcast pseudoultrasonic signals with location information. The self-developed hardware was connected in series to the speaker and consisted of the central control unit, a DDS module, and a bandpass filter. The main control module determines if the speaker is occupied. If it is not occupied, the DDS module generates a pseudoultrasonic signal with position coordinates and sends it out through the bandpass filter that is linked with the speaker. In Figure 5(a), we show the

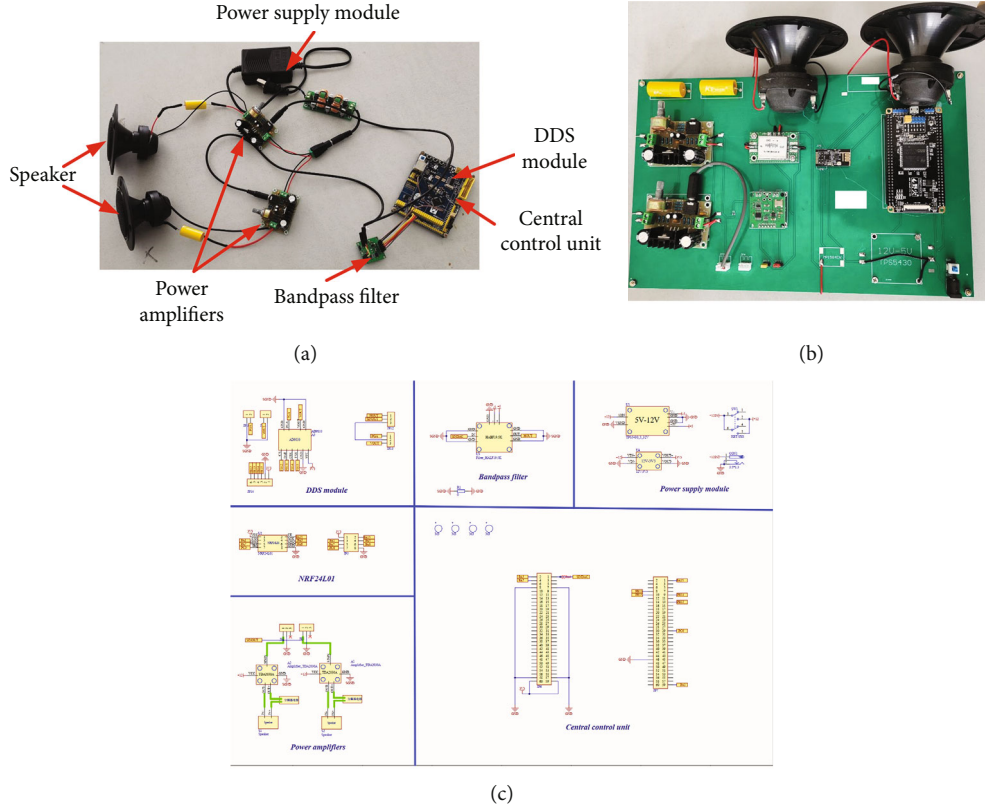


FIGURE 5: The physical connection. (a) The photo shows the connection. (b) Physical picture of the product. (c) PCB diagram.

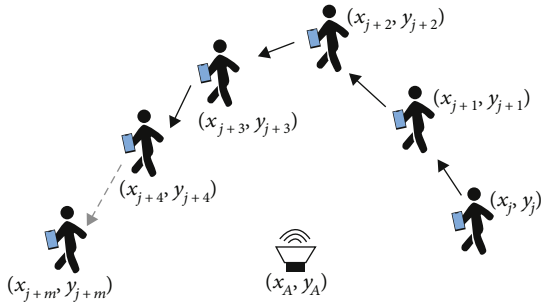


FIGURE 6: Schematic diagram of the acoustic source localization scene. The pedestrians are walking near the acoustic source with their smartphones.

connection diagram of each module. Further, we formed the preliminary product, as shown in Figure 5(b). Figure 5(c) shows the PCB (printed circuit board, PCB) diagram of our product

5.5.4. Acoustic Source Localization Algorithm Based on Hamming Distance. At present, most smartphones are equipped with dual microphones for noise reduction and other functions. For example, in the Android system, the dual microphones are strictly clock-synchronized when collecting audio signals. The pseudoultrasonic frequency range is 18 kHz-20 kHz with a wavelength of about 17 mm. Therefore, even if the distance between the two microphones of the mobile phone is short, the TDOA can be calculated using the time difference of the acoustic signal. The acoustic source

TABLE 2: Actions performed by volunteers.

	Holding (step)	Calling (step)	Swinging (step)	Pocket (step)
Walk	150	150	150	150
Run	150	150	150	150
Upstairs	150	150	150	150
Downstairs	150	150	150	150

TABLE 3: Pedestrian movement behavior recognition results.

	Walk	Run	Upstairs	Downstairs
Walk	<u>96.3%</u>	0.7%	0.8%	2.2%
Run	0.0%	<u>99.7%</u>	0.3%	0.0%
Upstairs	3.4%	0.8%	<u>92.1%</u>	3.7%
Downstairs	0.7%	0.6%	2%	<u>96.7%</u>

localization algorithm based on the Hamming distance consists of five steps: spatial gridding, generating a regional Hamming code, generating the acoustic source location Hamming code, matching the Hamming code, and calculating the acoustic source location.

(1) *Spatial Gridding.* When a pedestrian’s mobile phone receives a valid acoustic signal, it records the coordinates of the source in a grid and obtains several discrete grid points.

TABLE 4: Results of the recognition of the movement transition zone.

	Step 1 (walk)	Step 2 (upstairs)	Step 3 (upstairs)	Step 1 (walk)	Step 2 (downstairs)	Step 3 (downstairs)
SVM	96.4%	82.6%	89.4%	96.1%	90.7%	88.3%
SVM + D-S theory	97.1%	91.3%	92.5%	97.8%	95.2%	96.6%

(2) *Generating the Area Hamming Code.* When pedestrians are walking, the positioning area can be divided into two in a clockwise direction to determine the position and pose of the mobile phone. The left side is marked as 0, and the right side is marked as 1. Therefore, if a pedestrian walks M steps, we divide the positioning area into several small areas, each of which has a binary code of length M . The grid points of each small area also have the binary code of the small area.

(3) *Generating the Acoustic Source Location Hamming Code.* In this study, we binarize the TDOA value to increase the system's robustness, i.e., we use the vertical line of the mobile phone as the boundary. When moving in a clockwise direction, the acoustic source is denoted as 0 if it is on the left part of the vertical line (the TDOA has a negative value) and as 1 if it is on the right side of the vertical line (the TDOA has a positive value). Thus, if a pedestrian walks M steps, a binary code of length M is generated, indicating the area of the acoustic source.

(4) *Matching the Hamming Code.* We match the acoustic source location Hamming code with the regional Hamming code to determine the small area with the closest Hamming distance.

(5) *Calculating the Acoustic Source Position.* By matching the Hamming code, the acoustic source position can be reduced to a small area. However, since all grid points in the small area have the same binary code, the Hamming distance is the same for all the points in the small area. The average of the grid point positions is used as the estimated position of the acoustic source. However, the robustness of this method is too strong; thus, we use the methods in Refs. [30, 31] and use the proportional relationship between the Hamming distance and the actual distance for all grid points in the area. The source location Hamming code has different weights (the weight is inversely proportional to the Hamming distance). Therefore, the acoustic source position is determined by all grid points rather than the point with the smallest Hamming distance. We use a Gaussian function to obtain the weights of all grid points:

$$\omega_i = e^{-HD(T,D_i)/M\sigma_i}, \quad (14)$$

where $HD(T, D_i)$ is the Hamming distance between the acoustic source location Hamming code and the spatial grid point i ($i = 0, 1, 2 \dots p$), and σ is the weighting parameter. We normalize Equation (14) to obtain:

$$\bar{\omega}_i = \frac{\omega_i}{\sum_{i=1}^p \omega_i}. \quad (15)$$

The final acoustic source position $S(x_A, y_A)$ based on all grid point coordinates $p_i(x_i, y_i)$ as:

$$S = \sum_{i=1}^p \bar{\omega}_i \cdot p_i. \quad (16)$$

5.5.5. *Elimination of the Cumulative Error in PDR Positioning.* As shown in Figure 6, the mobile phone of the pedestrian receives the audio information and records the pedestrian's position as (\hat{x}_j, \hat{y}_j) . The acoustic source position (\hat{x}_A, \hat{y}_A) is recorded for M steps, and (\hat{x}_A, \hat{y}_A) is calculated at each step's position. PDR expresses the position (\hat{x}_k, \hat{y}_k) of any step k in the M steps as an expression of (\hat{x}_j, \hat{y}_j) , as shown in Equation (17).

$$\begin{bmatrix} \hat{x}_k \\ \hat{y}_k \end{bmatrix} = \begin{bmatrix} \hat{x}_j \\ \hat{y}_j \end{bmatrix} + \sum_{i=j}^M \hat{L}_i \cdot \begin{bmatrix} \sin(\theta_i) \\ \cos(\theta_i) \end{bmatrix}. \quad (17)$$

In the same way, the observed acoustic source position (\hat{x}_A, \hat{y}_A) is represented by (\hat{x}_j, \hat{y}_j) . If the accuracy of the observed acoustic source position meets the expectations, (\hat{x}_A, \hat{y}_A) is approximately equal to (x_A, y_A) . The position (\hat{x}_j, \hat{y}_j) is updated with the known acoustic source position information to obtain the coordinate (x_j', y_j') after the cumulative error has been eliminated. This method takes advantage of the high short-term accuracy of PDR positioning. There is already a cumulative error in the (\hat{x}_j, \hat{y}_j) position. However, due to the short running time of the PDR in M steps, the cumulative error is not large enough to affect the overall system; thus, it can be ignored.

Further, we analyze the PDR localization accuracy based on the correction of the acoustic source localization. We use Equations (10)–(12), and the latter is transformed as follows after performing the acoustic source localization correction:

$$\text{Error} = \begin{cases} \overline{\text{Error}} \cdot k, & (0 \leq n \leq k); \\ \overline{\text{Error}} \cdot (t - k) + \text{Error}_{\text{map1}}, & (k \leq n \leq t); \\ \overline{\text{Error}} \cdot (p - t) + \text{Error}_{\text{acoustic1}}, & (t \leq n \leq p); \\ \overline{\text{Error}} \cdot (r - p) + \text{Error}_{\text{map2}}, & (p \leq n \leq r); \\ \overline{\text{Error}} \cdot (N - r) + \text{Error}_{\text{acoustic2}}, & (r \leq n \leq N), \end{cases} \quad (18)$$

where $\text{Error}_{\text{acoustic1}}$ and $\text{Error}_{\text{acoustic2}}$ represent the positioning errors caused by the acoustic source positioning correction in the first and second step. After using the acoustic

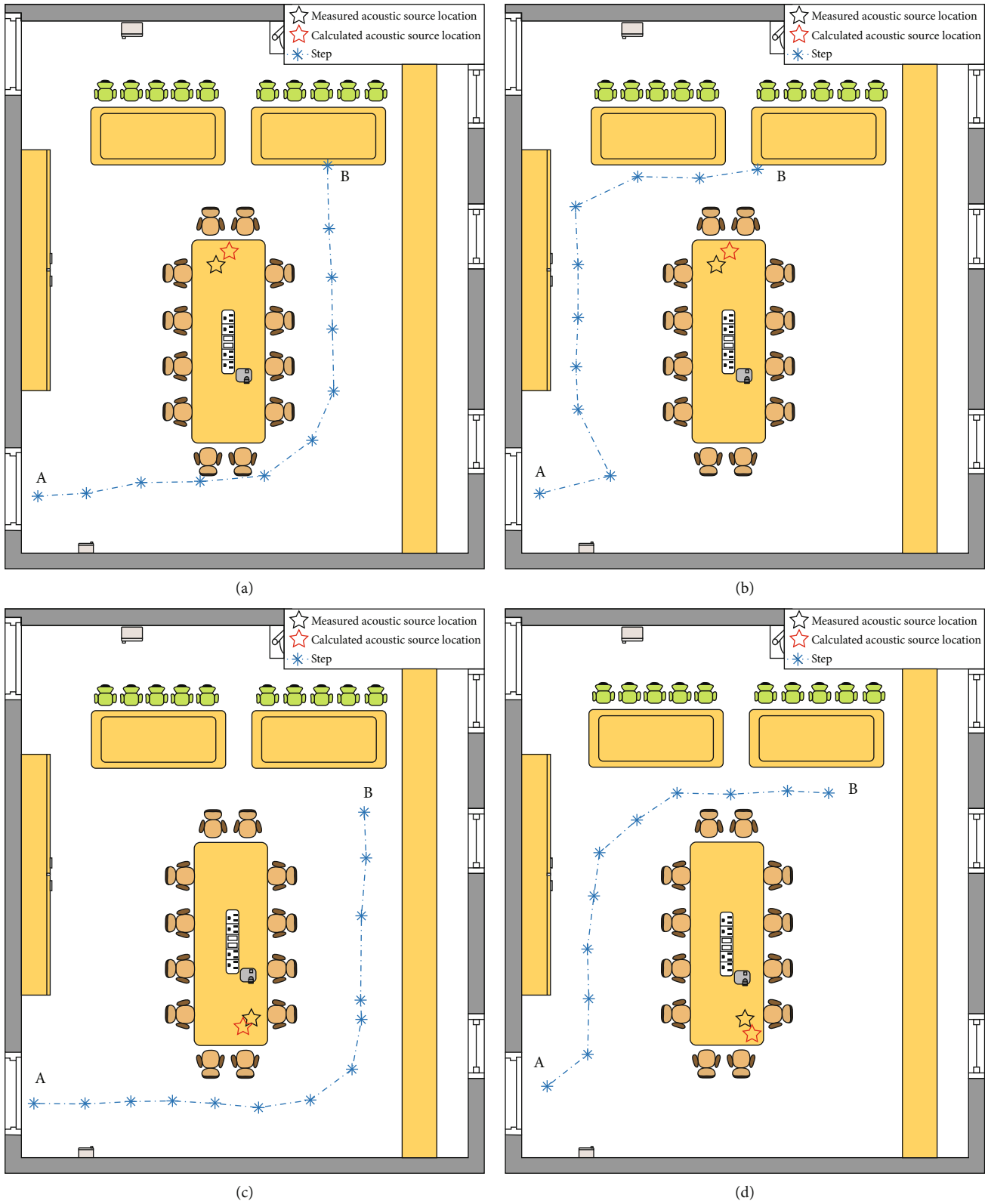


FIGURE 7: (d) Pedestrian trajectory in experiment. (a) In experiment 1; (b) in experiment 2; (c) in experiment 3; (d) in experiment 4.

source localization method, the global localization error is changed to a piecewise error. By segmenting the accumulated error, the large error can be decomposed into several smaller

errors, minimizing the global cumulative error. The results show that the proposed PDR positioning method based on acoustic source positioning correction provides relatively

TABLE 5: Results of the acoustic source localization experiment.

	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Actual acoustic source location (m)	(3.529, 4.353)	(3.529, 4.353)	(4.048, 6.904)	(4.048, 6.904)
Measured acoustic source location (m)	(3.75, 4.125)	(3.15, 4.3)	(3.9, 7.05)	(4.15, 7.15)
Root mean square error (m)	0.3175	0.3827	0.2079	0.2663

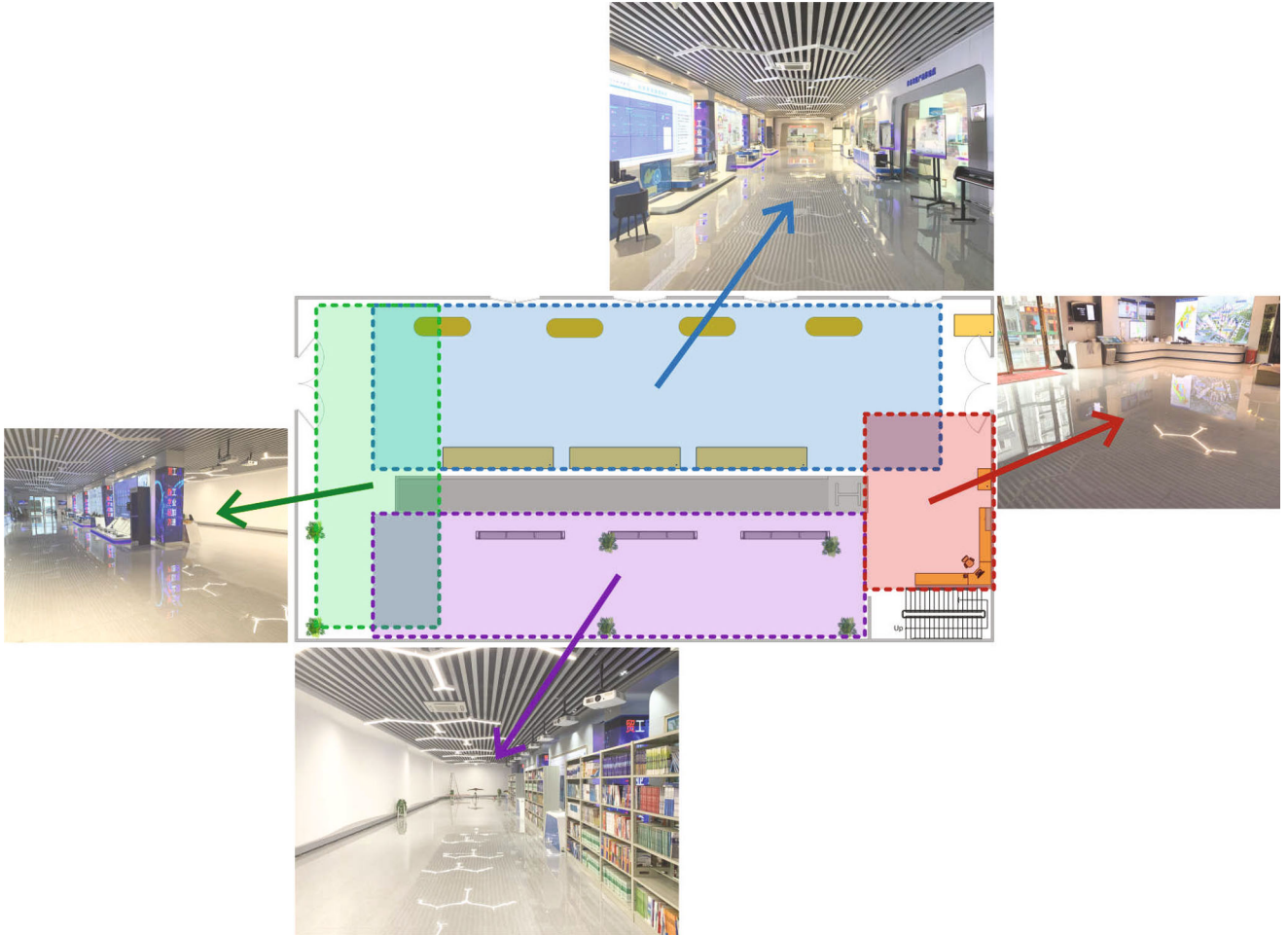


FIGURE 8: Schematic diagram of the stride estimation method experiment.

good positioning accuracy. The final positioning error is

$$\text{Error}(N) = \overline{\text{Error}} \cdot (N - r) + \text{Error}_{\text{acoustic2}}. \quad (19)$$

6. Experiment

6.1. Pedestrian Movement Behavior Recognition Experiment. Experiments were conducted to validate the proposed method. A Xiaomi 10 mobile phone with a sensor sampling frequency of 20 Hz is used. We recruited six male and female volunteers with different heights and weights to perform the actions listed in Table 2.

We used 70% of the feature data as the training set and 30% as the test set. The recognition accuracy is listed in following Table 3.

Further, we conducted experiments on the recognition of the movement transition zone, and we used two standard short stairs. We asked six volunteers to walk ten steps, walk up/down the stairs, and walk ten steps. We obtained 150 samples of walking upstairs and 150 samples of walking downstairs. The objective was to detect the transition zone of walking up/down the stairs (three steps). The result of using the SVM and the SVM + DS evidence theory is listed in Table 4.

Tables 3 and 4 show that the SVM provides reasonable accuracy for detecting single movements. However, the SVM method has lower accuracy than the SVM + D-S theory method for the recognition of the movement transition zone.

6.2. Acoustic Source Localization Experiment Based on Hamming Distance. We conducted an acoustic source

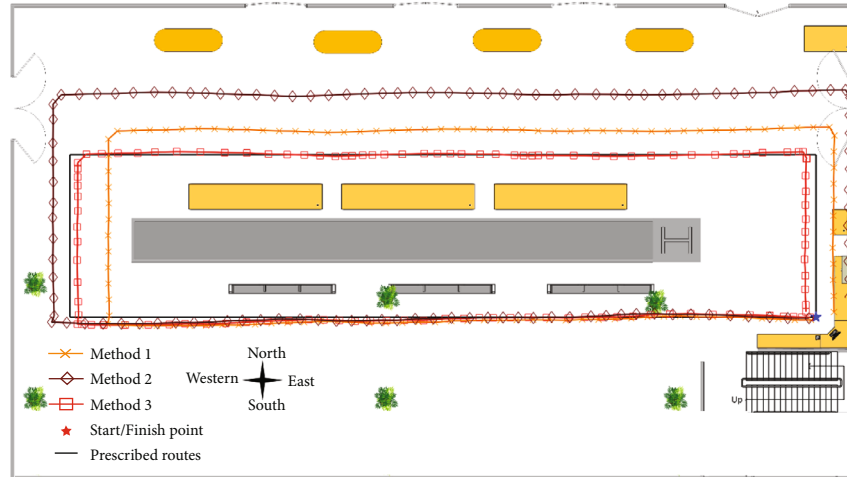


FIGURE 9: The trajectory in the stride estimation method experiment.

TABLE 6: Results of the stride estimation method experiment.

	Method 1	Method 2	Method 3
Coordinates	(1.06, -0.30)	(1.97,-1.2)	(-0.57, 0.30)
X-direction error (m)	1.06	1.97	0.57
Y-direction error (m)	0.30	1.2	0.30
Root mean square error (m)	1.10	2.30	0.67

localization experiment in a conference room with a size of $9.5 \text{ m} \times 8 \text{ m}$, as shown in Figure 7. A Huawei smart speaker was placed on the table in the middle of the conference room to simulate standard speaker equipment. We divided the four volunteers into two groups. The first group of students walked from point A to point B, and the second group of students walked from point B to point A. Four sets of data were collected. We used the electronic total station to obtain the coordinates of the speaker (as the black star in Figure 7), used PDR positioning to obtain the position of each step (as the blue dot in Figure 7), and use the acoustic source positioning method proposed in this paper to calculate the sound source position (as the red star in Figure 7). The results are shown in Figures 7(a)–7(d).

The actual and measured acoustic source locations and the root mean square errors are listed in Table 5.

The average error of the proposed acoustic source localization method is 0.2936 m , which met the expectations.

6.3. PDR Positioning Experiment. In this paper, an experiment is designed to verify the accuracy of the step estimation method based on pedestrian movement behavior recognition. The scene of this experiment is selected in the display hall of Guilin Smart Industrial Park, China. The scene diagram is shown in Figure 8. In the experiment, we stipulate volunteers to march along a prescribed route, in which volunteers first run forward to the west, and then walk north, east, and south in turn until they return to the starting point, with a total length of 108 m .

The experimental results are shown in Figure 9. Method 1 used a fixed stride length for PDR positioning (each stride

length was 0.7 m), method 2 used the method of Equation (6) to calculate the stride length (using fixed α and β), and method 3 was the method proposed in this paper. Let the coordinates of the starting point be $(0, 0)$, and the coordinates of the three methods and errors are shown in Table 6 after the volunteer walks one turn. The proposed method in this paper has the best positioning accuracy, and the trajectory of the volunteer walking best matches the prescribed trajectory.

A PDR positioning experiment was conducted in the seventh teaching building of Guilin University of Electronic Technology, Jinji Road. We asked the volunteers to travel along the route shown in Figure 10 from the starting point on the 3rd floor to the end of staircase B. The track was 70 m long, with two staircases (including a short staircase), and two acoustic sources were used.

The trajectory in the positioning experiment is shown in Figure 11.

The experiments verified that the proposed PDR and pedestrian movement behavior recognition eliminated the cumulative error of PDR positioning. As shown in Figure 11, when only PDR positioning was used and the running time was extended, the accumulated error of PDR positioning mistakenly located the pedestrians outside the building. Meanwhile, the root mean square error of positioning at the end point using only PDR is 2.15 m . Moreover, there were no prominent map landmarks such as stairs; therefore, the map matching algorithm could not be used to correct the PDR data. The acoustic source positioning eliminated the cumulative error, e.g., I and II. The correction performance is excellent for II, where the wrong route outside

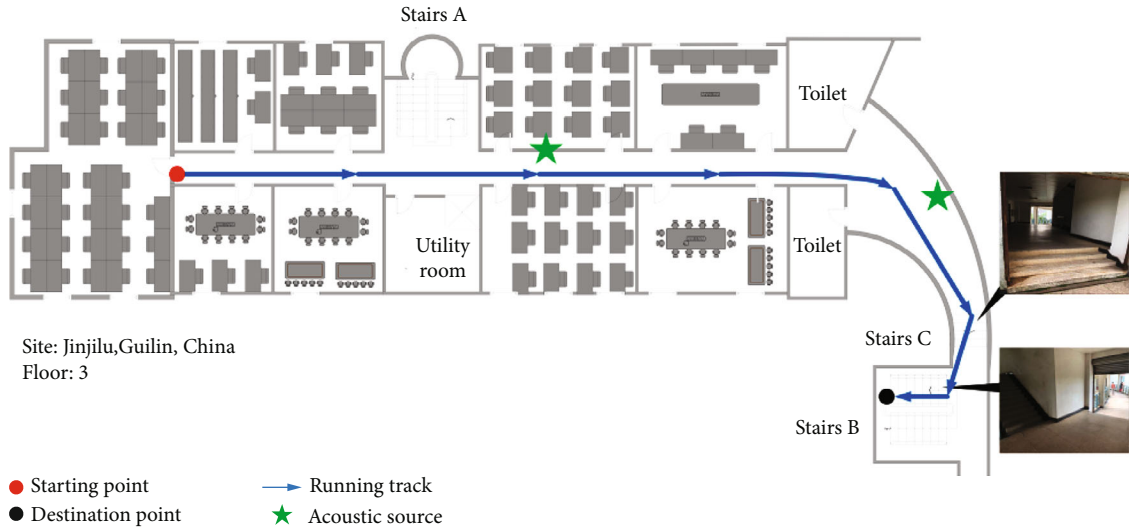


FIGURE 10: Schematic diagram of the positioning experiment.

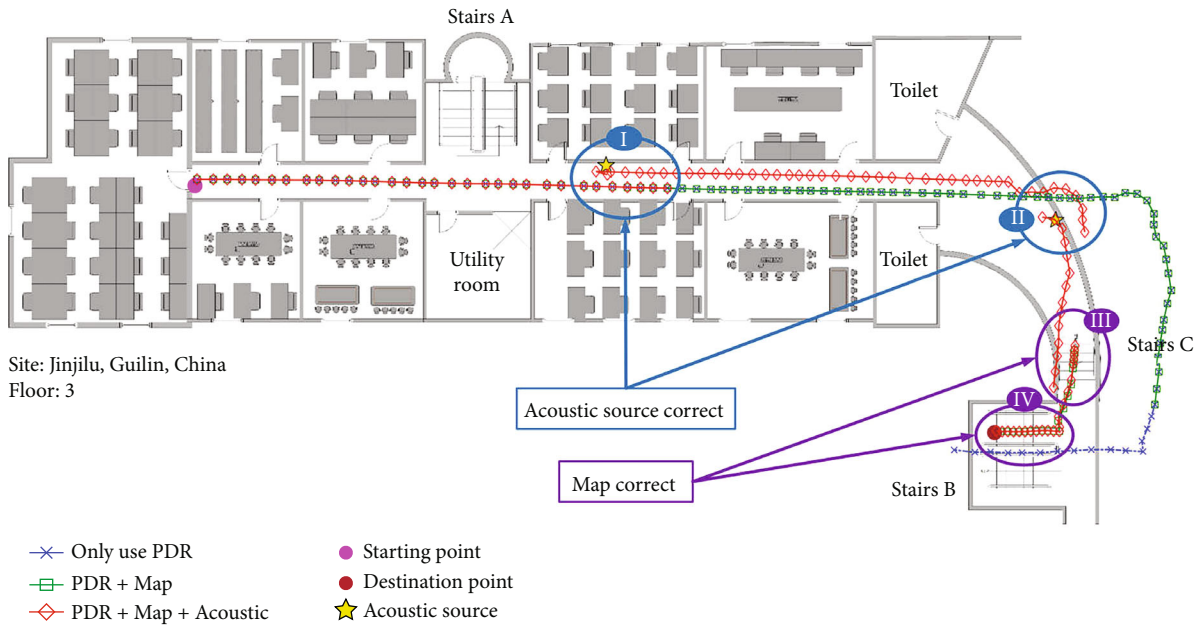


FIGURE 11: The trajectory in the positioning experiment.

the building is corrected. In addition, without the pedestrian movement behavior recognition method, the movement behavior states, such as going up and down the stairs, cannot be identified, resulting in a wrong trajectory. However, using the map matching methods (III, IV) results in correct stride length estimation and the elimination of the accumulated error of PDR positioning.

7. Conclusions and Future Work

This paper proposed a new PDR positioning method that optimizes the PDR positioning algorithm by identifying the pedestrian movement behavior recognition. The experiments showed that the recognition accuracy exceeded 96%, which met the expectations. The acoustic source positioning

method eliminated the cumulative error of PDR positioning, resulting in a positioning error of 0.3 m, which met the requirements. The proposed method shows excellent potential for positioning applications. In the future, we will investigate the use of this method in more complex scenarios.

Data Availability

The Pedestrian Movement Behavior Recognition data and Positioning data used to support the findings of this study have not been made available because this paper is funded by the Guangxi Innovation-Driven Development Project (Science and Technology Major Project of Guangxi No. AA18118039). The grant is still in the research phase and

all research data is currently restricted to disclosure within the project team.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research work was funded by the Guangxi Innovation-Driven Development Project (Science and Technology Major Project of Guangxi No. AA18118039), the Guangxi Keypoint Research and Invention Program (No. GuiKe AB17292058), the Guangxi Science and Technology Plan Project (No. AD18281020 and No. AD18281044), and the Director Fund project of the Key Laboratory of Cognitive Radio and Information Processing of Ministry of Education (No. CRKL190104).

References

- [1] F. Van Diggelen, *A-Gps: Assisted Gps, Gnss, and Sbas*, Artech house, 2009.
- [2] "Stage 2 functional specification of user equipment (UE) positioning in E-UTRAN," 3GPP TS 36.305 version 10.0.0.
- [3] F. Zafari and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2568–2599, 2019.
- [4] M. Murata, D. Ahmetovic, D. Sato, H. Takagi, K. M. Kitani, and C. Asakawa, "Smartphone-based indoor localization for blind navigation across building complexes," in *2018 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 1–10, Athens, Greece, 2018.
- [5] A. R. Jiménez and F. Seco, "Comparing Decawave and Bespoon UWB location systems: indoor/outdoor performance analysis," in *2016 international conference on indoor positioning and indoor navigation (IPIN)*, pp. 1–8, Alcalá de Henares, Spain, 2016.
- [6] A. Poulou, J. Kim, and D. S. Han, "A sensor fusion framework for indoor localization using smartphone sensors and Wi-Fi RSSI measurements," *Applied Sciences*, vol. 9, no. 20, p. 4379, 2019.
- [7] R. W. Levi and T. Judd, *Dead Reckoning Navigational System Using Accelerometer to Measure Foot Impacts*, US Patent 5,583,776, 1996.
- [8] S. Zhuang, "Real-Time Indoor Location Tracking in Construction Site Using BLE Beacon Trilateration," 2020.
- [9] Z. Hao, J. Dang, W. Cai, and Y. Duan, "A multi-floor location method based on multi-sensor and WiFi fingerprint fusion," *IEEE Access*, vol. 8, pp. 223765–223781, 2020.
- [10] M. Wang, Z. Chen, Z. Zhou, J. Fu, and H. Qiu, "Analysis of the applicability of dilution of precision in the base station configuration optimization of ultrawideband indoor TDOA positioning system," *IEEE Access*, vol. 8, pp. 225076–225087, 2020.
- [11] G. Shi, Y. Zou, Y. Jin, Y. Zheng, and W. J. Li, "Multi-category human motion recognition based on MEMS inertial sensing data," in *2009 4th IEEE international conference on Nano/micro engineered and molecular systems*, pp. 489–493, 2009.
- [12] G. Li, L. Zhang, F. Lin, M. Chen, and Z. Wang, "SAIloc: a novel acoustic single array system for indoor localization," in *2017 9th international conference on wireless communications and signal processing (WCSP)*, pp. 1–6, Nanjing, China, 2017.
- [13] T. M. T. Dinh, N. S. Duong, and K. Sandrasegaran, "Smartphone-based indoor positioning using BLE iBeacon and reliable lightweight fingerprint map," *IEEE Sensors Journal*, vol. 20, no. 17, pp. 10283–10294, 2020.
- [14] Z. Li, L. Zhao, C. Qin, and Y. Wang, "WiFi/PDR integrated navigation with robustly constrained Kalman filter," *Measurement Science and Technology*, vol. 31, no. 8, article 084002, 2020.
- [15] Y. Li, Y. Zhuang, P. Zhang, H. Lan, X. Niu, and N. el-Sheimy, "An improved inertial/wifi/magnetic fusion structure for indoor navigation," *Information Fusion*, vol. 34, pp. 101–119, 2017.
- [16] Y. Yao, L. Pan, W. Fen, X. Xu, X. Liang, and X. Xu, "A robust step detection and stride length estimation for pedestrian dead reckoning using a smartphone," *IEEE Sensors Journal*, vol. 20, no. 17, pp. 9685–9697, 2020.
- [17] B. Wang, X. Liu, B. Yu, R. Jia, and X. Gan, "Pedestrian dead reckoning based on motion mode recognition using a smartphone," *Sensors*, vol. 18, no. 6, p. 1811, 2018.
- [18] L. Zheng, X. Zhan, X. Zhang, S. Wang, and W. Yuan, "Heading estimation for multimode pedestrian dead reckoning," *IEEE Sensors Journal*, vol. 20, no. 15, pp. 8731–8739, 2020.
- [19] W. Kang and Y. Han, "SmartPDR: smartphone-based pedestrian dead reckoning for indoor localization," *IEEE Sensors Journal*, vol. 15, no. 5, pp. 2906–2916, 2015.
- [20] M. Alzantot and M. Youssef, "UPTIME: ubiquitous pedestrian tracking using mobile phones," in *2012 IEEE wireless communications and networking conference (WCNC)*, pp. 3204–3209, Sydney, Australia, 2012.
- [21] D.-K. Cho, M. Mun, U. Lee, W. J. Kaiser, and M. Gerla, "Autogait: a mobile platform that accurately estimates the distance walked," in *2010 IEEE international conference on pervasive computing and communications (PerCom)*, pp. 116–124, Mannheim, Germany, 2010.
- [22] V. Renaudin, M. Susi, and G. Lachapelle, "Step length estimation using handheld inertial sensors," *Sensors*, vol. 12, no. 7, pp. 8507–8525, 2012.
- [23] A. C. Dİrican and S. Aksoy, "Step counting using smartphone accelerometer and fast fourier transform," *Sigma Journal of Engineering and Natural Sciences*, vol. 8, pp. 175–182, 2017.
- [24] N. Mohssen, R. Momtaz, H. Aly, and M. Youssef, "It's the human that matters: accurate user orientation estimation for mobile computing applications," in *Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering)*, pp. 70–79, 2014.
- [25] A. Mannini and A. M. Sabatini, "Machine learning methods for classifying human physical activity from on-body accelerometers," *Sensors*, vol. 10, no. 2, pp. 1154–1175, 2010.
- [26] K. Altun, B. Barshan, and O. Tunçel, "Comparative study on classifying human activities with miniature inertial and magnetic sensors," *Pattern Recognition*, vol. 43, no. 10, pp. 3605–3620, 2010.
- [27] F. Höflinger, R. Zhang, J. Hoppe et al., "Acoustic self-calibrating system for indoor smartphone tracking (assist)," in *2012 international conference on indoor positioning and indoor navigation (IPIN)*, pp. 1–9, Sydney, Australia, 2012.

- [28] K. Liu, X. Liu, and X. Li, "Guoguo: enabling fine-grained smartphone localization via acoustic anchors," *IEEE Transactions on Mobile Computing*, vol. 15, no. 5, pp. 1144–1156, 2016.
- [29] L. Zhang, M. Chen, X. Wang, and Z. Wang, "TOA estimation of chirp signal in dense multipath environment for low-cost acoustic ranging," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 2, pp. 355–367, 2018.
- [30] N. Jin, X. Zhou, Z. Wang, Y. Liu, and L. Wang, "Robust sequence-based localization in acoustic sensor networks," in *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 3809–3813, Calgary, AB, Canada, 2018.
- [31] N. Jin, X. Zhou, C. Lin et al., "ThunderLoc: smartphone-based crowdsensing for thunder localization," in *2018 15th annual IEEE international conference on sensing, communication, and networking (SECON)*, pp. 1-2, HongKong, China, 2018.
- [32] Z. Zhou, N. Duan, T. Li et al., "Human moving behavior recognition based on the DS evidence theory and SVM utilizing smartphone," in *2020 IEEE Intl Conf on dependable, autonomous and secure computing, Intl Conf on pervasive intelligence and computing, Intl Conf on cloud and big data computing, Intl Conf on cyber science and technology congress (DASC/PiCom/CBDCCom/nnnnnSciTech)*, pp. 277–283, Calgary, AB, Canada, 2020.
- [33] C. Yu, N. el-Sheimy, H. Lan, and Z. Liu, "Map-based indoor pedestrian navigation using an auxiliary particle filter," *Micro-machines*, vol. 8, no. 7, p. 225, 2017.

Research Article

Intelligent IoT-Based Cross-Border e-Commerce Supply Chain Performance Optimization

Lei Xia¹ and Sitong Liu² 

¹*School of Economics and Management, Xi'an Aeronautical University, Xi'an, 710077 Shaanxi, China*

²*School of Management, Guilin University of Aerospace Technology, Guilin, 541004 Guangxi, China*

Correspondence should be addressed to Sitong Liu; liusitong@guat.edu.cn

Received 25 March 2021; Revised 17 May 2021; Accepted 8 June 2021; Published 22 June 2021

Academic Editor: Peng Yu

Copyright © 2021 Lei Xia and Sitong 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.

Internet of Things (IoT) technology can benefit automated production, agriculture, intelligent autonomous driving, and many other industries by using billions of smart devices. As a good example, intelligent IoT and 5G are the main source of information acquisition and play an important role in the multiobjective optimization process of the supply chain. This paper develops an optimal management and coordination method to improve the performance of cross-border e-commerce supply chain by using IoT tracking technique and multiobjective decision-making. The numerical results justify that our proposed scheme has high internal consistency, with the Cronbach's alpha factor of each dimension of the optimized decision model all greater than 80%.

1. Introduction

Over the past decade, a number of emerging technologies have changed the evolving path of our life. 5G Internet of Things (IoT) is a global network based on standard communication protocol, which intends to collect observations from physical world and implement practical applications [1–3]. It is predicted that by 2025, 75.44 billion devices worldwide will be connected to the Internet of Things. The Internet of Things technology is the next important technology to realize the interconnection of everything in the world. In the ten years from 2020 to 2030, IoT devices will grow from 75 billion to more than 100 billion [4]. They have the characteristics of large scale, higher speed, more mode, and high data quality and heterogeneity [5]. As the main driving force of IoT, 5G is supposed to have extended coverage, higher throughput, less waiting time, and large-scale bandwidth connecting density [2], laying a way to connect hundreds of billion sensors between networks. As a result, uniform and heterogeneous sensor networks can be connected to large-scale sensing devices, making great contributions to human beings provide advanced and intelligent services.

In recent years, with the gradual increase in the popularity of e-commerce and consumer consumption, domestic products have become only a part of consumers' buying wishes. More and more people have been attracted by the shopping list of cross-border e-commerce platforms. High-quality overseas products have driven the development of cross-border e-commerce (CBE) industry.

The cross-border e-commerce supply chain, with the help 5G IoT networks, becomes a value-added chain. The value of materials in the supply chain is increased due to processing, packaging, transportation, and tracking. Related companies in the logistics chain, information chain, and capital chain are getting benefit too. A complete e-commerce supply chain should include suppliers, manufacturers, distributors, retailers, and consumers, under the umbrella of an intelligent IoT platform.

The market changes rapidly, so the requirements passed to the supply chain will also change with the trends in the market, and the requirements of the supply chain are different at different stages of development. For example, in the initial stage, it is necessary to improve quality, reduce costs, and make products more competitive in the terminal; the

initial stage requires the supply chain to fully release production capacity to ensure market supply; the market differentiation stage requires the supply chain to become more flexible and more and more flexible. It can adapt to the production of multiple varieties and small batches. Different from the company's business plan, the focus of supply chain performance should also be updated at any time.

In the literature, Kawa believes that e-commerce is one of the most dynamic and important sectors of economic activity [6]. The latest trend in the e-commerce market is cross-border trade, based on the sale of various products to customers in other countries. Kawa has conducted research on several problems in e-commerce, such as long time and high cost. The purpose of the research is to put forward the concept of a cross-border e-commerce integrator, which will solve the supply chain performance problems of cross-border e-commerce. For this reason, Kawa proposed the concept of a cross-border e-commerce integrator whose main task is to integrate the entire supply chain. Thanks to the economies of scale obtained by integrating the packages of many e-shops, integrators can achieve lower delivery costs in international transportation, return and exchange products more effectively, and better serve customers from different countries. This research method lacks theoretical and factual support and is not suitable for promotion in reality.

Prompanyo and Wang, instead, stated in [7] that the main purpose of his research is to verify the stability of the multi-scale measurement of perceived value in the electronic loyalty model and explain the interrelationship between perceived values. In the context of cross-border e-commerce, they objectively evaluate the tradeoff of satisfaction and e-loyalty.

Prompanyo and Wang use confirmatory factor analysis and Bootstrap structural equation model to evaluate research hypotheses. The research results show that the multidimensional scale of perceived value can describe and explain the influence of perceived value on e-loyalty model. In this study, the overall perceived value positively and indirectly affects e-loyalty through the mediation of network satisfaction. The process of this method is more complicated and error prone.

Turkulainen and Swink, alternatively, endeavored to evaluate the contextualized operation effects of cross-border e-commerce supply chain personnel (involving logistics and supply management) [8]. On the basis of the classic contingency theory, the research conducted a detailed contextual analysis of the performance impact of internal supply chain personnel participating in innovation activities. Specifically, Turkulainen and Swink believe that the impact of corporate product innovation activities on operational effectiveness is regulated by the organization's technical background, the level of integration of operating suppliers, and the interaction between these two factors. The research results show that, with the assistance of supplier integration, internal supply chain personnel in an organization can be a source or channel of valuable innovation-related knowledge, especially in a high-tech environment. In addition, since the complementary effects are highly dependent on the supply chain performance dimension, the results further support the argument for decomposing performance. Turkulainen and Swink discussed the effectiveness of supply chain personnel participat-

ing in product innovation activities, by explaining the link mechanism between enterprise product innovation activities and performance, and established the basic relevance of supply chain personnel participating in product innovation activities. The research application cost is relatively high, and it is not suitable for popularization in practice.

Different from the existing work, this paper has the following contributions.

- (1) It proposes the overall optimization inventory strategy of the supply chain through inventory adjustment, in a bid to improve the management of cross-border e-commerce supply chain
- (2) It suggested a novel coordination model to optimize the performance of the cross-border e-commerce supply chain
- (3) It takes into account the optimization of the efficiency of cross-border e-commerce supply chain

The remainder of the paper is organized as follows. Section 2 first proposes and formulates the CBE supply chain optimal management and coordination problem. Section 3 presents the detailed design of optimization and coordination schemes. Section 4 finally gives the numerical results of decision-making and coordination mechanism, followed by Section 5 to conclude the paper.

2. Formulation of Optimal Management and Coordination Methods for CBE Supply Chain

2.1. Big Data Mining Framework Based on 5G Internet of Things. Big data mining is widely used in sensing regions and object processors. Compared to traditional data, there are multiple versions of big data generated in the 5G Internet of Everything, which requires more real-time analysis. Figure 1 illustrates that the 5G IoT data mining involves 3 steps, i.e., data collection, transmission, and then preprocessing. The data from fixed and mobile sensing areas may have distinct quality levels, and thus, preprocessing of the raw data needs to be done in the object processor.

2.2. Optimize Decision-Making through Inventory. Supply chain management should include supply chain planning, coordination and control, information flow and capital flow between participating organizations, and parts in the supply chain. Its purpose is to optimize the speed and certainty of all related programs and maximize all related programs. The net added value of the process improves the operational efficiency and efficiency of the organization [9]. An important part of the supply chain system is the research and analysis of the supply chain storage system. Generally speaking, in the sales process of CBE, in order to meet the various needs of its customers in time, avoid shortages and delayed delivery. When things happen, it is necessary to hold a certain cost inventory [10]. Also, during the procurement and production period, in order not to interrupt the production process and maintain the continuity of production, cross-border e-commerce must have a certain amount of raw material

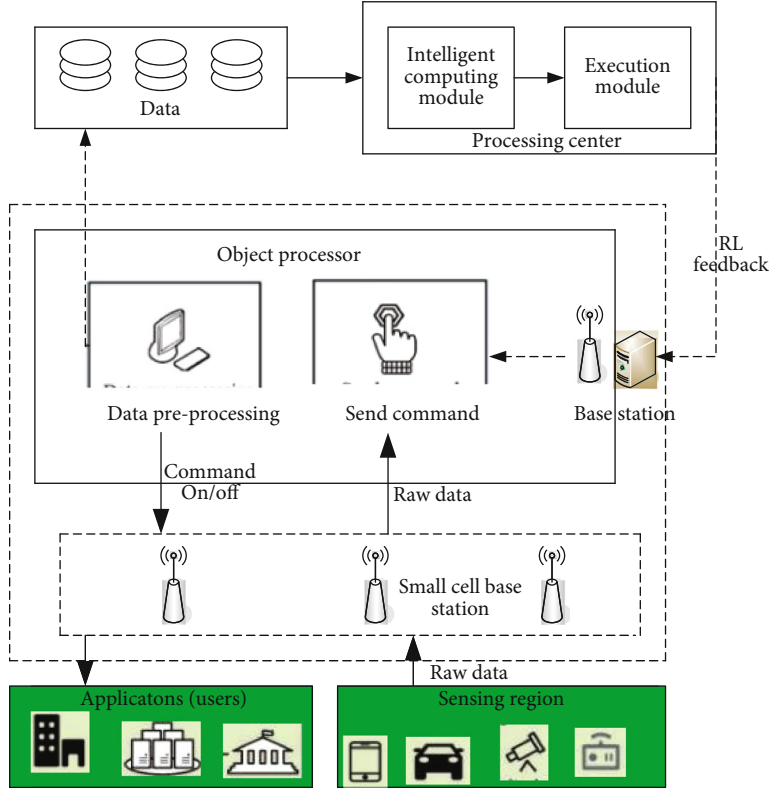


FIGURE 1: Big data mining framework.

inventory [11]. Due to many reasons, the amount of cross-border e-commerce inventory is often uncertain. If the inventory reserve is too small, it will not be able to meet the needs of customers. The excess reserve will not only occupy a large amount of funds, but also waste a lot of production area and inventory area, resulting in inventory backlog [12, 13]. Therefore, it is necessary to study how to determine a supply chain-oriented overall optimization inventory strategy through coordination. This problem must be comprehensively studied in order to achieve overall planning. It is usually necessary for sales, supply, production, finance, and other departments to work together to formulate a reasonable inventory quota, often revised and adjusted as the situation changes.

Assume that both the supplier and the demander adopt a periodic inspection strategy, that is, (t, S) replenishment strategy [14]. This strategy refers to checking the inventory at regular intervals and placing an order to replenish the existing inventory to the maximum inventory level S . If the inventory is I during the check, the order is $Q = S - I$. At the same time, it is assumed that the supply and demand parties adopt a one-to-many replenishment model, that is, the replenishment cycle of the distributor is an integer multiple of the retailer's replenishment cycle. In this way, it is possible to popularize and apply the inventory decision theory under the one-to-one replenishment model (that is, assuming that the replenishment cycle of both parties is the same) [15]. Suppose the entire planning period H is divided into n equal parts, the length of each equal part is $1 = H/n$, and

the replenishment time point $T_m = (i - 1)H/n$ at the beginning of the m^{th} period. According to the assumption, the retailer's demand per unit time is $f(t)$, so the demand from T_m at the beginning of the m^{th} period to time t is $F_m(t) = \sum_{x=T_m}^t f(x)$. Because the demand $f(x)$ obeys a normal distribution, it can be seen that $F_m(t)$ also obeys a normal distribution [16]. In practice, the customers that retailers face are general ones. In order to more accurately and effectively reflect the changes of customer demand over time, the unit time length of demand must be short. The time is regarded as a continuous variable in the hypothesis, in a bid to respond to customer demand in a timely manner [6]. As $F_m(t) \approx \int_{T_m}^t f(x)dx$, the mean and variance of the demand are then calculated by $\mu_{F_m(t)} = \int_{T_m}^t \mu(x)dx$, $\sigma_{F_m(t)}^2 = \int_{T_m}^t \sigma^2(x)dx$, respectively. When the retailer's demand $F_m(t)$ from the beginning of the m^{th} period T_m to time t is lower than the basic inventory level $S_{k,m}$ at the beginning of the period, it is considered that the retailer will have inventory held at the time t of the m^{th} period, and its holding inventory is $S_{k,m} - F_m(t)$. When the retailer's demand $F_m(t)$ from the beginning of the m^{th} period T_m to time t is higher than the basic inventory level $S_{k,m}$ at the beginning of the period, it is considered that the retailer is out of stock at the time t of the m^{th} period, and its out-of-stock quantity is $F_m(t) - S_{k,m}$ [9]. Both the out-of-stock quantity and the holding inventory are random processes. The average holding inventory of the retailer at time t in the m^{th} period is as follows:

$$E(S_{k,m} - F_m(t))^+ = \int_0^{s_{k,m}} (S_{k,m} - x) \varphi_{F_m(t)}(x) dx. \quad (1)$$

Out of stock quantity is as follows:

$$E(F_m(t) - s_{k,m})^+ = \int_{s_{k,m}}^{+\infty} (x - s_{k,m}) \varphi_{F_m(t)}(x) dx. \quad (2)$$

The average inventory cost of the retailer in the m^{th} period is as follows:

$$\begin{aligned} MC_{k,m} = & c_k + \int_{T_m}^{T_{m+1}} h_k E(S_{k,m} - F_m(t))^+ dt \\ & + \int_{T_m}^{T_{m+1}} p_k E(F_m(t) - S_{k,m})^+ dt. \end{aligned} \quad (3)$$

Substituting formula (1) and formula (2) into formula (3), a new relational expression can be obtained:

$$\begin{aligned} MC_{k,m} = & c_k + h_k \int_0^{S_{k,m}} (S_{k,m} - x) \int_{(m-1)1}^{ml} \varphi_{F_m(t)}(x) dt dx \\ & + p_k \int_{S_{k,m}}^{+\infty} (x - S_{k,m}) \int_{(m-1)1}^{ml} \varphi_{F_m(t)}(x) dt dx. \end{aligned} \quad (4)$$

From this, the retailer's inventory cost throughout the planning period has the following relationship:

$$\begin{aligned} MC_k = & nc_k + h_k \sum_{m=1}^n \int_0^{S_{k,m}} (S_{k,m} - x) \int_{(m-1)1}^{ml} \varphi_{F_m(t)}(x) dt dx \\ & + p_k \sum_{m=1}^n \int_{S_{k,m}}^{+\infty} (x - S_{k,m}) \int_{(m-1)1}^{ml} \varphi_{F_m(t)}(x) dt dx. \end{aligned} \quad (5)$$

The first term of the formula refers to the retailer's total order cost during the planning period, the second term refers to the retailer's total holding cost during the planning period, and the third term refers to the retailer's total out-of-stock penalty cost within [17, 18]. In order to optimize the retailer's basic inventory level (formula (5)), the retailer's inventory cost can be derived with respect to $S_{k,m}$, and the following formula can be obtained:

$$\begin{aligned} \frac{\partial MC_k}{\partial S_{k,m}} = & h_k \int_0^{S_{k,m}} \int_{(m-1)1}^{ml} \varphi_{F_m(t)}(x) dt dx \\ & - p_k \int_{S_{k,m}}^{+\infty} \int_{(m-1)1}^{ml} \varphi_{F_m(t)}(x) dt dx \\ = & (h_r + p_r) \int_{(m-1)1}^{ml} \int_0^{S_{k,m}} \varphi_{F_m(t)}(x) dt dx - p_k 1. \end{aligned} \quad (6)$$

Standardize the random variable $F_m(t)$ that obeys the normal distribution, let $F_m(t) - \mu_{F_m(t)} / \sigma_{F_m(t)} = Z$, and we know that the random variable X also obeys the standard normal distribution [19]. Therefore, it can be seen that the

values of random variable $F_m(t)$ and random variable Z have the following relationship:

$$\frac{x - \mu_{F_m(t)}}{\sigma_{F_m(t)}} = Z. \quad (7)$$

Substitute formula (7) into formula (6) to obtain a new relationship:

$$\frac{\partial MC_k}{\partial S_{k,m}} = (h_r + p_r) \int_{(m-1)1}^{ml} \phi\left(\frac{S_{k,m} - \mu_{F_m(t)}}{\sigma_{F_m(t)}}\right) dt - p_k 1. \quad (8)$$

Let formula (8) be equal to zero, and the optimal basic inventory level formula for retailers can be obtained as follows:

$$\int_{(m-1)1}^{ml} \phi\left(\frac{S_{k,m} - \mu_{F_m(t)}}{\sigma_{F_m(t)}}\right) dt = \frac{p_k 1}{h_k + p_k}. \quad (9)$$

2.3. Adopt a Coordination Model of Cross-Border e-Commerce Supply Chain Contracts. Since consumers' shopping experience is directly linked to the logistics service level of third-party logistics companies, it will have a certain impact on the business performance of cross-border e-commerce. Therefore, in order to encourage third-party logistics companies to make progress in the level of logistics services, cross-border e-commerce companies are willing to share a certain proportion of the cost for the third-party logistics companies in exchange for their own profits [20]. At the beginning of the period, a cross-border e-commerce company orders a certain product quantity Q from a supplier based on market forecasts, pays the order fee C_1 and the inventory management cost C_2 of the self-built bonded warehouse, and determines the price of a certain product p according to the market and cost: third-party logistics company. After understanding the market situation and a certain product type of the cross-border e-commerce company, when consumers place an order with the cross-border e-commerce company, the logistics service fee P_C of the order is charged and the logistics service of the logistics service level s is provided [21]. When the third-party logistics company incurs the cost of improving the logistics service level, the cross-border e-commerce company will voluntarily pay the part of the cost that accounts for $(1 - \theta)(0 < \theta < 1)$ [22]. Based on this, the profit functions of cross-border e-commerce companies and third-party logistics companies are, respectively, as follows:

$$\begin{aligned} R_1 = & (p - p_C)S(Q, s) - (c_1 + c_2)Q - c_e I(Q, s) \\ & - c_u L(Q, s) - (1 - \theta)g(s), \end{aligned} \quad (10)$$

$$R_2 = (p_C - c_C) \times S(Q, s) - \theta g(s). \quad (11)$$

The warehouse in the bonded zone can be built into an imported logistics distribution center. According to the law, those warehouses can store foreign goods and enjoy the bonded policy. Therefore, foreign companies can be

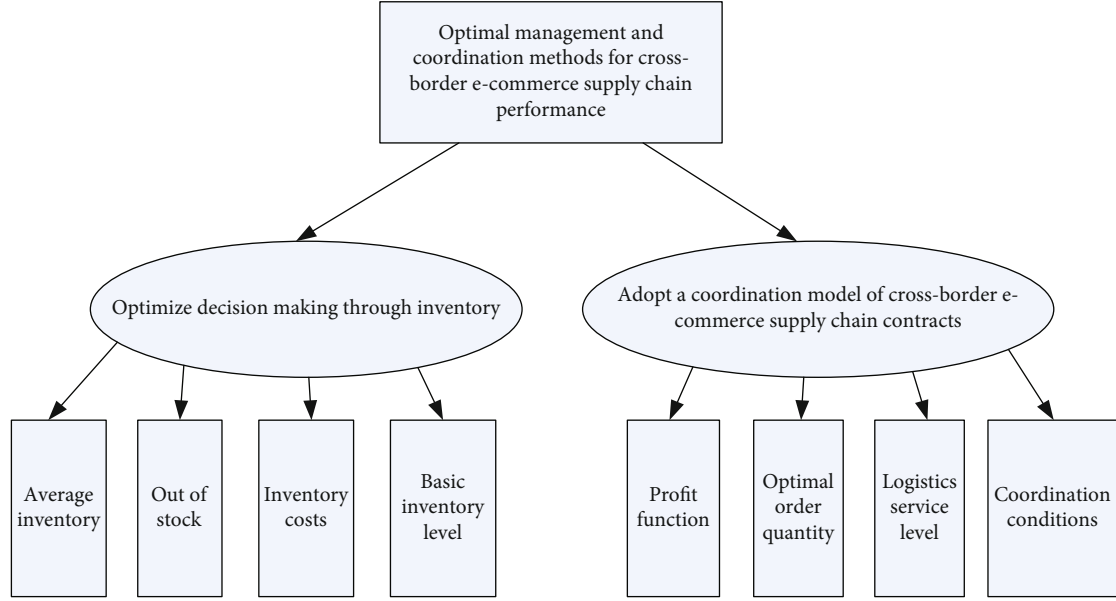


FIGURE 2: Part of the technical flow chart of this method.

organized to directly store the imported raw materials required by domestic enterprises in the bonded warehouse. When the company needs to import raw materials, purchasing at the warehouse in the bonded zone can form an overseas goods market within the country by build the bonded warehouse into an import logistics distribution center. Under the coordination of competition alliances, cross-border e-commerce companies and third-party logistics companies are still in the role of decentralized decision-making. The starting point is still to maximize their own interests. Cross-border e-commerce companies determine the optimal order quantity of goods, and third-party logistics company determines the optimal logistics service level in this situation [23]. The optimal order quantity of a product of a cross-border e-commerce company satisfies the following formula:

$$Q_1 = F^{-1} \left(\frac{p - p_C - c_1 - c_2 + c_u}{p - p_C + c_e + c_u} \right) - bp + ds. \quad (12)$$

The logistics service level s_2 of the third-party logistics company satisfies the following formula:

$$s_2 = \frac{d[(1 - \varphi)(p - p_C + c_e + c_u) + p_C - c_C]}{\theta k} \times F(Q). \quad (13)$$

Substitute formula (13) into formula (12) to get a new relationship:

$$s_2 = \frac{d[(1 - \varphi)(p - p_C + c_e + c_u) + p_C - c_C]}{\theta k} \times \frac{p - p_C - c_1 - c_2 + c_u}{p - p_C - c_e - c_u}. \quad (14)$$

If a cross-border e-commerce company logistics service supply chain under a cost-sharing and revenue-sharing com-

ination contract wants to achieve a coordinated state, it should meet $Q_1 = Q^*$, $s_2 = s^*$, and the relevant formulas should be combined to obtain the conditions for achieving supply chain performance coordination under the combined contract for

$$p_C = (\theta + \varphi - 1)p + (1 - \theta)c_C + \theta c_e + \theta c_u. \quad (15)$$

In the cross-border e-commerce company's import cross-border e-commerce logistics service supply chain system based on the combination contract, the decision variables of the optimal order quantity Q and logistics service level s of a certain commodity are made by the cross-border e-commerce company and the third-party logistics company, respectively, decision [24]. To achieve the coordination of system decision-making, whether for Q or s , the profit function of each member under the combined contract can be transformed into the affine function of $R_0(Q)$ and $R_0(s)$ under centralized decision-making, that is, $R_{1(Q)} = aR_0(Q) + b$ (where a and b are independent of Q constants).

The method part of this article uses the above methods to study the performance optimization management decision-making and coordination mechanism of cross-border e-commerce supply chain based on multiobjective optimization. Research is based on multiobjective optimization, cross-border e-commerce supply chain, performance management optimization, coordination mechanism, etc. The specific process is shown in Figure 2.

3. Experiment Study on Management and Coordination of CBE Supply Chain

3.1. Design a CBE Supply Chain Management Optimization Plan. Products and services are the core of each company's market competition. Optimizing supply chain performance

can effectively reduce the operating costs of products in terms of manufacturing, transportation, and delivery, improve operating efficiency and market response speed, and enhance the company's service level. Supply chain optimization needs to start from the overall point of view to realize the harmonious symbiosis of upstream and downstream enterprises, reduce mutual internal friction, and improve the efficiency of the entire supply chain [25]. Optimize the general supply chain to achieve a fairer and more reasonable distribution of the benefits of each enterprise in the supply chain, better serve the end customers, and realize the value of the entire supply chain. Through optimization, an efficient and streamlined supply chain must be established to improve the competitiveness of cross-border e-commerce companies in the market; and links that are not suitable for improving the supply chain efficiency of cross-border e-commerce companies and waste corporate resources should be adjusted [26].

(1) Principles of supply chain performance optimization

- (1) *Fully Identify and Meet Consumer Needs.* Supply chain management realizes the process of products from satisfying consumer needs to product design, production, processing, packaging, logistics warehousing, and distribution [27]. Products that meet consumer needs are the value transfer process of a successful supply chain, and the follow-up part is a necessary process to realize this value transfer.
- (2) *Shorten the Product Supply Chain Cycle.* The product supply chain cycle is the entire process of product design, production, and sales to consumers. If the supply cycle is too long, it will be at a disadvantage in a rapidly changing market [28]. Shortening the product supply chain cycle can shorten the time from product development, procurement to customer feedback, and make companies more competitive.
- (3) *Shorten the Testing Time of New Products.* For cross-border e-commerce platforms, from product selection, promotion, to consumer feedback data analysis, to feedback to suppliers, bulk purchase, production and logistics transportation to mass sales, a certain timeliness is required [29]. It is possible to use methods such as manual intervention and social marketing inside and outside the platform to quickly implement new product evaluation and promotion and obtain customer feedback as soon as possible to respond to market changes quickly.
- (4) *Reduce Supply Chain Operating Costs.* Supply chain operating cost is the necessary cost from the production of a company's products to the sales process [30]. Operating costs will ultimately be reflected in product sales prices and profits, and reducing supply chain operating costs will help companies better control profits, give prod-

uct sales prices more room for adjustment, and improve product competitiveness in the market.

- (5) We are not afraid of more scenes, but the scenes should not be specifically customized. If it is too customized and cannot be copied, the cost will be high and the business will fail. We have to respond to the needs of customers and use general multiscenario solutions to digest customer requirements and resolve their problems

(2) Design optimization steps

According to the general idea and principle of optimization, in order to achieve better results, the following steps need to be optimized.

First of all, there are more or less problems in each link in the supply chain. Enterprises need to start from the appearance of the problem, find the essence of the problem, and analyze the connection between the problem and the problem in each link [31].

Secondly, starting from the essence of the problem, combining the value and development of the entire chain, give the most suitable but not the optimal solution, so that the problems in each link can be comprehensively solved, the cost can be controlled, and the process can be executed. The result is satisfactory [32].

Then, after the plan comes out, there must be a dedicated person in charge to supervise the implementation of the plan and implement feedback during the implementation of the plan. If the expected effect cannot be achieved, the plan needs to be optimized. During the implementation of the optimization plan, the plan needs to be tracked in time. For the relevant implementation situation, it is necessary to analyze and solve the new factors that affect the resolution of the problem in a timely manner [33].

Finally, after the implementation of the entire program is completed, the corresponding evaluation system is used to evaluate and score the effect of the program to ensure that the company has both input and output for each program implementation. Do a good job in the production and sorting of text materials and data materials and record and sort the problems and points that can be improved during the implementation of the plan and expand the company's experience database for further improvements in the future optimization plan.

3.2. Realize Cross-Border e-Commerce Supply Chain Performance Optimization. Sort out the business processes of existing companies importing cross-border e-commerce, sort out suitable business processes and system processes, and rely on advanced information technology to establish an operating mechanism and an efficient IT management system that can quickly support new business expansion. At the same time, through the reconstruction of the existing business process, the customs clearance business is service-oriented to provide support for upstream businesses, and at the same time, the business that is not related to customs clearance is separated from the system, such as "accepting orders, docking with warehouse systems, and docking with

TABLE 1: Cross-border e-commerce business process.

Step	1	2	3		
Process node	Order node	Transfer node	Check joints	Parcel production node	
Step	5	6	1	9	
Process node	Outbound declaration and loading node	Loading declaration node	Delivery distribution node	Delivery node	Sign node

e-commerce platforms.” Establish a reasonable business system, maintain the purity of production, and reduce the complexity of the production system. Establish an order management system to conduct unified, standardized, and precise control of cross-border e-commerce and general trade orders. Establish a three-dimensional production, operation, and maintenance monitoring system for product service level, timeliness assessment, and KPI assessment to create an efficient service platform.

Receive sales orders from major platforms. When receiving orders, the system needs to perform data verification. After the verification is passed, the corresponding platform will be notified whether the order is successfully received. When receiving the order, the relevant data that need to be verified for legality are sorted out so that the system can promptly remind. After the order is successfully received, the system needs to determine whether the order meets the production requirements. It needs to check the goods, merchants, and inventory. When the order is successfully transferred, it needs to inform the platform that the order has been successfully received and the production is in progress; then, the customs and the customs clearance and declaration work of the national inspection system include the issuance of declaration work orders, the issuance of three declaration instructions (orders, waybills, and payment orders), three declaration receipt processing, list declaration instructions issuance and list declaration receipt processing, and customs clearance. After the completion, the package production can be arranged; the order will be delivered to the warehouse for production after customs clearance is completed, and the packaging of the entire order will be completed. The package production node mainly includes order distribution, wave generation, product picking, order picking, and order packaging. After the production of the order package is completed, before loading the corresponding order on the truck, you need to complete the customs declaration activities of the order from the warehouse. After the declaration is completed, you can load the corresponding order on the truck and leave the bonded area; when the order is loaded, the final declaration is required, that is, the customs inspection order is about to leave the bonded area to complete the order delivery process. When the order loading declaration is completed, you can leave the bonded area and deliver it to the domestic logistics distributor; when the order loading declaration is completed and the truck leaves the bonded area, the order needs to be handed over to the corresponding distributor to start domestic delivery. The entire link needs system and physical handover with logistics distributors; when the distributor receives the package, it completes domestic distribution according to the scheduled route, and the logistics dis-

tributor needs to return real-time logistics distribution information to the e-commerce platform and the logistics tracking system of the enterprise. After the parcel is delivered to the customer, the customer will sign for the receipt. At this time, the customer’s receipt information will be sent back to the e-commerce platform and the enterprise’s logistics tracking system to complete the overall business process of cross-border e-commerce. Draw this cross-border e-commerce business process into a table, as shown in Table 1.

Through the abovementioned business sorting and process standardization, the main process of the entire cross-border e-commerce import business is clarified, and then, by combining this process, relevant cross-border e-commerce companies can design the overall business system to meet the smooth development of the abovementioned business.

3.3. Realize Cross-Border e-Commerce Logistics Management Optimization

(1) Innovative logistics distribution model

At present, there are two innovative distribution models for cross-border e-commerce logistics, namely, logistics outsourcing and the establishment of multinational e-commerce logistics cooperation alliances. Traditional cross-border e-commerce logistics and distribution include three types: the overseas direct mail mode that sells goods to consumers by means of postal parcels, commercial express, or logistics lines, and this method is generally expensive, and by means of bonded areas and free trade zones. Stocking mode: this mode has certain requirements for inventory estimation; self-built overseas warehousing mode: but the overseas warehousing construction cost is high and the cycle is long.

- (1) For many small and medium-sized cross-border e-commerce companies that cannot afford high logistics costs, they can use logistics outsourcing; the first typical logistics outsourcing method is third-party logistics, assisted by self-operated distribution models, and remote areas are selected postal services; the second is the fourth-party logistics, which provides logistics planning, consulting, and supply chain management activities for both parties and third parties. It integrates resources from domestic and foreign merchants and logistics companies and integrates them on the website through Internet information technology. For consumers to choose the best solution, the fourth-party logistics e-commerce platform can also carry out unified customs declaration

TABLE 2: Major global cross-border e-commerce platforms.

Nation	North America	Europe	South America
Electronic business platform	Amazon, Bonanza, Cratejoy, eBay, Etsy, Jet, Newegg, Reverb, Walmart, Wayfair, Wish, Zibbet	Allegro, Asos, Cdiscount, Cel. DaWanda.Emag. Flubit, Fnac. Fruugo, Game. Mobile.de. Okazii, OnBuy, PriceMinister. Real.de, Tesco. Zalando, Otto	Americanas, Casas Bahia, Dafiti, Extra, Linio, Mercado Libre. Submarino
Number of platforms	12	18	7
Nation	Asia	Africa	Oceania
Electronic business platform	Alibaba, AliExpress, Flipkart, GittiGidiyor, HipVan, JD, Kaola, Lazada, Qoo10, Lotte. Shopee, Snapdeal, Souq, TaoBao, tmall. vip	Jumia, Kilimall, Konga	Iconic.MyDeal
Number of platforms	16	3	2

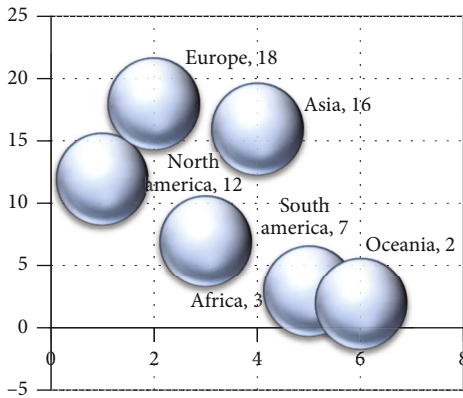


FIGURE 3: Major global cross-border e-commerce platforms

for cross-border e-commerce companies, pay taxes on their behalf, and conduct centralized commodity inspection reviews

- (2) Establish a multinational e-commerce logistics cooperation alliance. The logistics cooperation alliance is a logistics model in addition to self-operated logistics, third-party and fourth-party logistics. It refers to two or more companies that are formed on the basis of benefit-sharing with other companies in order to make up for their own logistics
- (3) *Competitiveness. Joint Action and Partnership.* China's first cross-border e-commerce cooperation alliance was established in Lanzhou in January 2016. After its establishment, the alliance integrated international railway freight trains and the e-commerce resources of various countries to form a smooth and efficient logistics channel, which also gave cross-border e-commerce the future development direction of the logistics of commercial and import enterprises has pointed the way.
- (2) Improve the logistics guarantee mechanism

First, invest in the construction of a reverse logistics information tracking system. An efficient and complete

TABLE 3: Development of cross-border e-commerce in the past five years.

Age	Transaction amount (unit: trillion)	Growth rate
2015	5.47	30.24%
2016	6.72	22.85%
2017	8.06	19.94%
2018	9.12	13.15%
2019	10.31	19.44%

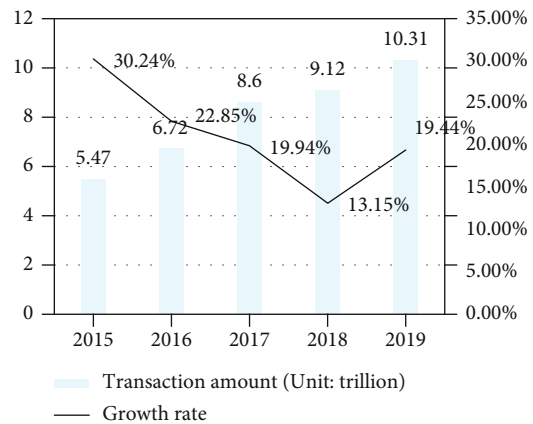


FIGURE 4: Development of cross-border e-commerce in the past five years.

reverse logistics information system should be able to integrate the information of each node member of the supply chain. Through this system, cross-border e-commerce import companies can know the user ID of reverse logistics the reason for return and the cost; logistics companies can find the location, time, and quantity of the returned goods; foreign suppliers can trace the origin to the production place and supply channels of the product for continuous attention; users can estimate the cost and processing time of the return. The reverse logistics system requires enterprises to arrange and combine all information to form a database and apply it to the operation of the enterprise. Enterprises can use this

TABLE 4: Distribution of cross-border e-commerce companies.

Area	Beijing	Shanghai	Guangdong	Zhejiang	Jiangsu	Other areas	Nationwide
Number of cross-border e-commerce companies	238	326	4229	1746	703	6041	13,283
Percentage	1.79%	2.45%	31.84%	13.14%	5.29%	45.48%	1

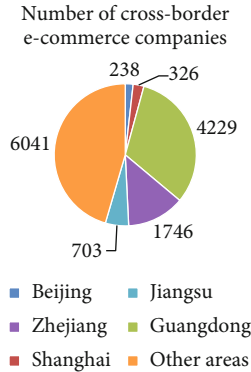


FIGURE 5: Distribution of cross-border e-commerce companies.

system to deal with consumer returns and exchanges in time based on database information.

Second, improve the online purchase process. E-commerce companies or platforms should first provide product information as much as possible in the online store. Even if the customer has placed an order and paid, if you want to cancel the order and refund, the customer service staff should deal with it as soon as possible after asking the original situation to reduce unnecessary delivery. Costs reduce the return rate. At the same time, companies should collect, analyze, and summarize relevant information about returned goods, find out possible problems with returned goods, and optimize management methods to reduce the return rate in response to these problems.

4. Decision-Making and Coordination Mechanism Analysis

4.1. Macroenvironment Analysis of Cross-Border e-Commerce Business

(1) Political environment

With the rapid development of globalization, cooperation between countries is becoming more and more close. China's "one belt, one road" policy is launched on this basis, which is also an important project for our country to go abroad. Since the one belt, one road one belt, one road has been signed. We have signed the development agreement with more than 30 countries along the way. The main form of cooperation is e-commerce cooperation. e-commerce has also become an important part of the "one belt and one way" development cooperation, playing a key role in the economic and trade cooperation between our country and other countries. Another relevant research shows that the main form of global economic and trade cooperation is e-commerce. Currently, there are 58 major e-commerce plat-

forms in the world. Cross-border e-commerce is an important part of it. The details are shown in Table 2 and Figure 3.

As can be seen from the chart, Europe has the most e-commerce platforms and the most mature development, with a total of 18 e-commerce platforms, followed by China, with a total of 16 e-commerce platforms, of which Taobao is the most famous; North America is ranked third in the number of e-commerce platforms. The most famous of which is the Amazon e-commerce platform. Although China's e-commerce platform and cross-border e-commerce business have started slowly, they have developed rapidly and often set new transaction records at the end of each year.

(2) Economic environment

As the world's second-largest economy, although China's economic growth has slowed in recent years, its economic growth still ranks first in the world. The form of China's imports and exports has risen steadily. With the rapid development of China's cross-border e-commerce industry, more and more SMEs and individuals have joined the cross-border e-commerce industry, which has accelerated the development of the industry. At the same time, the rise of domestic production costs has also accelerated the transformation of the industry, and the mode of foreign trade has gradually shifted to the online. Under this situation, cross-border e-commerce platforms are developing rapidly, and at the same time, the competition among cross-border e-commerce companies has increased unprecedentedly. According to China Customs statistics, in 2019, China's total import and export trade amounted to 31.86 billion yuan, an average annual growth rate of 21%. According to the latest statistics from iResearch, the scale of China's cross-border e-commerce reached 94.1 billion yuan in 2019, a year-on-year increase of 22.7% and a growth rate of 29.41%. Based on the calculation of these previous data, Ali believes that the transaction volume of China's cross-border e-commerce will rise to 13.15 trillion in 2020, and the growth rate will reach 1.35 trillion 28.52%, a growth rate of 39.1%. The future cross-border development of e-commerce has a huge market. Draw the specific situation into a chart, as shown in Table 3 and Figure 4.

The report shows that in terms of cross-border e-commerce model structure, cross-border e-commerce B2B represents 83 companies. In 2018, China's cross-border e-commerce accounted for 2%, and cross-border e-commerce transaction B2C accounted for 16.8%. The B2B model accounts for more than 80% of cross-border e-commerce models and has been the dominant business model for many years. The business model of cross-border e-commerce B2B lies in disintermediation, allowing owners and product names to communicate directly and dealing with producers

TABLE 5: Cost comparison before and after implementation of cross-border e-commerce supply chain optimization plan.

Project	Before implementation	After implementation
Commodity purchase cost	82.00	82.00
International freight+port fees	7.04	7.04
Tariff	7.05	/
Value-added tax	14.78	/
Comprehensive cross-border e-commerce tax	/	13.31
Packaging supplies	0.80	0.80
Cross-border warehouse distribution service fee	/	15.00
Guangzhou warehouse storage	3.80	/
Domestic express delivery	12.13	/
Sales loss	1.76	1.76
Profit	8.23	8.23
Selling price	137.56	125.74

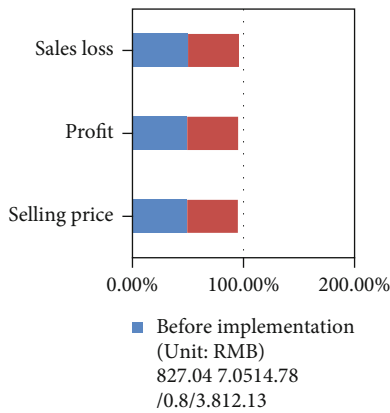


FIGURE 6: Cost comparison before and after implementation of cross-border e-commerce supply chain optimization plan.

and trademarks. More and more B2C cross-border e-commerce platforms have been established.

Factories and consumers are directly connected to many Internet networks, reducing trade links and eliminating information asymmetry in the form of B2B2C. The B2C model is more flexible than traditional foreign trade and other forms.

(3) Social environment

In today’s economic globalization, the rapid development of the Internet has accelerated the exchange of the world. The globalization of economy and commodity has made the dependence between countries closer and has brought into play the advantages of various countries and promoted their consumption and development. With the smooth flow of consumption channels in various countries, products with personal characteristics and preferences are more favored by people. Cross-border e-commerce provides them with shopping channels, enabling them to purchase their favorite products through cross-border e-commerce. Therefore, cross-border e-commerce has entered the era of take-off. With the development of network culture, consumers have

more ways to understand foreign brand culture and at the same time purchase a large number of foreign goods through cross-border shopping, and this process continues to promote the spread and reputation of foreign brands in China. To promote the rapid growth of imported e-commerce, cross-border e-commerce promotes a new round of consumption upgrades in China and becomes a new driving force for the sustained growth of China’s national economy. For enterprises, it is necessary to seize this opportunity and make good use of the Internet platform to achieve a win-win situation for enterprises and consumers. The rapid development of cross-border e-commerce has benefited from the rapid development of Internet technology. The Internet is a “catalyst” for the development of cross-border e-commerce. The rapid development of the Internet, big data, and information technology has shortened. With the development of mobile information technology and the change of people’s past communication modes, consumers can obtain market information more efficiently and quickly. The development of informatization in the warehousing and logistics industries ensures that consumers can obtain dynamic information about commodities. Real-time control and the improvement of diversified payment methods and security technology guarantee consumer safety. The development of science and technology has provided a solid foundation and guarantee for cross-border e-commerce. According to the search results of Tianyan Check, there are a total of 13,283 formal enterprises specializing in cross-border e-commerce across the country, which are located in major cities or small and medium-sized cities in China. The specific conditions are shown in Table 4 and Figure 5.

According to statistics, most of China’s formal enterprises specializing in cross-border e-commerce are located in Zhejiang Province, where Alibaba was founded. The rest are mostly located in developed areas such as Beijing and Shanghai. There are few cross-border e-commerce-related companies in the western region. It can be seen that in order to optimize the performance of cross-border e-commerce supply chain management, decision-making, and coordination, at least the company should be located in a convenient transportation and economically developed area.

TABLE 6: Results of scale reliability analysis.

Dimension	Cronbach's alpha	Number of items	Dimension
Supply chain management capabilities	0.803	7	Supply chain management capabilities
Online marketing	0.821	5	Online marketing
Inventory management	0.842	5	Inventory management
Brand management	0.887	5	Brand management
Representation	0.846	9	Representation

4.2. *Evaluation and Analysis of Cost Optimization of Cross-Border e-Commerce Supply Chain Optimization Plan.* The cost composition of cross-border imported goods include procurement costs, international logistics, import customs clearance and taxation, warehousing, packaging, domestic logistics, and sales loss. Take a well-known cross-border e-commerce company as an example. Before the optimization, the company's business model was to import through general trade, pay tariffs+value-added tax normally, store goods in Guangzhou's own warehouses, and deliver goods from Guangzhou warehouses to all parts of the country. After the implementation of the supply chain optimization plan, the company's business model was adjusted to stock up through the cross-border e-commerce comprehensive experimental zone supervision library, and the import of goods paid cross-border e-commerce comprehensive tax, and the third-party logistics company was responsible for the implementation of "warehouse and distribution integration" warehousing and delivery of goods. The cost composition comparison before and after the implementation of the company's cross-border e-commerce supply chain optimization plan is shown in Table 5 and Figure 6.

Under the same conditions, the sales price of the company's cross-border supply chain optimization plan is reduced by 8.59% compared to before implementation. The total tax on cross-border e-commerce is equal to 69% of the tax. The domestic adjusted tax rate is calculated based on the general business model and the increased tax rate. After the adjustment, the cost reduction rate is greatly reduced. The supply of the latest e-commerce service is 14% higher than the previous domestic price, but after replacing the storage fee of the Guangzhou warehouse, the distribution cost of the warehouse has been reduced by about 11%. At the same time, storage and distribution links are entirely third parties, which greatly reduce the company's hidden management costs for storage and distribution links.

4.3. *Reliability Test Analysis of Cross-Border e-Commerce Supply Chain Optimization Decision Model.* The degree of reliability plays a critical role in measuring the consistency level of the results achieved by a specific approach. In particular, the credibility of internal cohesion mainly relies on Cronbach's alpha factor to test the consistency of responses to all elements on the same scale. Table 6 and Figure 7 demonstrate our analysis using the above criterion.

The reliability analysis part is to take the reliability factor (Cronbach β) according to the problems of each level to understand the consistency of the model rating. In this paper, the Cronbach β value of each scale factor of the optimized

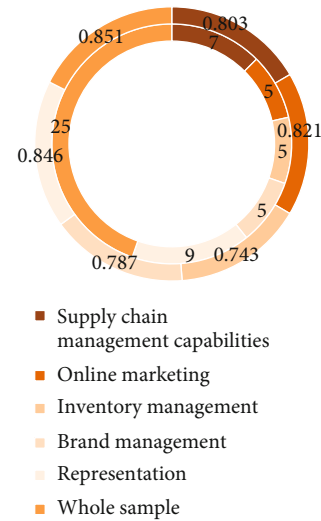


FIGURE 7: Results of scale reliability analysis.

decision model is greater than 0.80. From a statistical point of view, the reliability coefficient of any test or scale is above 0.80, indicating that the internal consistency of the test or scale is good. This shows that the optimized decision-making coordination mechanism has high internal consistency.

5. Conclusion

The advent of 5G and IoT has changed people's lifestyles and promoted the creation of new business models. As a product of the Internet era, cross-border e-commerce is a concrete manifestation of "Internet + foreign trade". In recent years, it has developed rapidly. There are more and more traditional foreign trade companies, especially small and medium foreign trade companies, and they have begun to transform into a cross-border e-commerce model. In this process, companies urgently need to optimize and upgrade the original traditional international supply chain system to cover the needs of the global cross-border e-commerce industry.

In the early stage of the research, this paper puts forward the optimization management and coordination method of cross-border e-commerce supply chain performance. To determine the overall optimized inventory strategy oriented by the supply chain through inventory coordination has become an important means of cross-border e-commerce supply chain performance management; adopting the coordination model of cross-border e-commerce supply chain contracts, in order to encourage third-party logistics companies to improve in terms of logistics service levels, cross-border e-commerce

companies are willing to share a certain proportion of the cost for third-party logistics companies in exchange for their own profits. This paper also proposes to design a cross-border e-commerce supply chain performance management optimization plan, to achieve cross-border e-commerce supply chain performance optimization, and to achieve cross-border e-commerce logistics management optimization.

Justified by the numerical results, this article conducts comprehensive analysis. We conclude that although China's e-commerce platform and cross-border e-commerce business have started slowly, they have developed rapidly, and the transaction volume record is often refreshed at the end of each year; the performance of the cross-border e-commerce supply chain must be improved. To optimize management decision-making and coordination, at least try to locate the company in an area with convenient transportation and developed economy. Finally, it is clear that our decision model pushes the Cronbach's alpha factor of all dimensions greater than 0.80, indicating a high internal consistency with coordination mechanism.

Data Availability

The data used in this article is the company's private data.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the National social science major project in 2017 "Research on reform and innovation of supervision system of state owned enterprises" (17ZDA087).

References

- [1] A. Whitmore, A. Agarwal, and L. Da Xu, "The Internet of Things — a survey of topics and trends," *Info. Systems Frontiers*, vol. 17, no. 2, pp. 261–274, 2015.
- [2] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: a comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 18, pp. 1617–1655, 2017.
- [3] L. Atzori, A. Iera, and G. Morabito, "Understanding the Internet of Things: definition, potentials, and societal role of a fast evolving paradigm," *Ad Hoc Networks*, vol. 56, pp. 122–140, 2017.
- [4] S. H. Shah and I. Yaqoob, "A survey: Internet of Things (IOT) technologies, applications and challenges," in *2016 IEEE Smart Energy Grid Engineering (SEGE)*, pp. 381–385, 2016.
- [5] X. Du, M. Guizani, Y. Xiao, and H.-H. Chen, "Routing-driven elliptic curve cryptography based key management scheme for heterogeneous sensor networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 3, pp. 1223–1229, 2009.
- [6] A. Kawa and W. Zdrenka, "Conception of integrator in cross-border E-commerce," *Logforum*, vol. 12, no. 121, pp. 63–73, 2016.
- [7] M. Prompanyo and L. Wang, "A validation of the multidimensional perceived value in the model of E-loyalty towards Sino-Thai cross-border E-commerce based on China's customers," *Journal of Business Research-Turk*, vol. 12, no. 2, pp. 1014–1022, 2020.
- [8] V. Turkulainen and M. L. Swink, "Supply chain personnel as knowledge resources for innovation—a contingency view," *Journal of Supply Chain Management*, vol. 53, no. 3, pp. 41–59, 2017.
- [9] B. Lu and H. Wang, "Research on the competitive strategy of cross-border E-commerce comprehensive pilot area based on the spatial competition," *Scientific Programming*, vol. 2016, no. 1, 2016.
- [10] M. A. A. Aqlan, "Research on the status quo and countermeasures of cross-border E-commerce development in Arab countries," *Open Journal of Business and Management*, vol. 8, no. 4, pp. 1536–1542, 2020.
- [11] S. Ji, X. Wang, W. Zhao, and D. Guo, "An application of a three-stage XGBoost-based model to sales forecasting of a cross-border E-commerce enterprise," *Mathematical Problems in Engineering*, vol. 2019, no. 2, 2019.
- [12] S. Ji, "Research on personalized recommendation algorithm of cross-border e-commerce under large data background," *Italian Journal of Pure and Applied Mathematics*, vol. 41, pp. 358–368, 2019.
- [13] J. Mou, Y. Cui, and K. Kurcz, "Trust, risk and alternative website quality in B-buyer acceptance of cross-border E-commerce," *Journal of Global Information Management*, vol. 28, no. 1, pp. 167–188, 2020.
- [14] L. Shi, S. Liu, and S. Petrovi, "Cryptanalysis of a pseudorandom generator for cross-border E-commerce," *Ingénierie des systèmes d'information*, vol. 24, no. 4, pp. 361–365, 2019.
- [15] P. Wang, "On the development of cross-border E-commerce and the transformation of foreign trade model," *Modern Economy*, vol. 9, no. 10, pp. 1665–1671, 2018.
- [16] K. Gao, Y. Huo, and Y. Liu, "An investment decision-making research on cross-border e-commerce overseas warehouse based on real option," *ICIC Express Letters*, vol. 11, no. 5, pp. 1005–1010, 2017.
- [17] A. J. Lin, E. Y. Li, and S. Y. Lee, "Dysfunctional customer behavior in cross-border E-commerce: a justice-affect-behavior model," *Journal of Electronic Commerce Research*, vol. 19, no. 1, pp. 36–54, 2018.
- [18] Y. Fang, "Current situation, obstacles and solutions to China's cross-border E-commerce," *Open Journal of Social Sciences*, vol. 5, no. 10, pp. 343–351, 2017.
- [19] Y. K. Wang, "Model for evaluating the logistics service quality of cross-border E-commerce enterprises with intuitionistic fuzzy information," *Journal of Computational and Theoretical Nanoscience*, vol. 14, no. 2, pp. 1136–1139, 2017.
- [20] L. Yang, J. Chen, H. Zhang, H. Jiang, S. A. Vorobyov, and D. T. Ngo, "Cooperative wireless multicast: performance analysis and time allocation," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 7, pp. 5810–5819, 2016.
- [21] E. Chancey, J. L. M. Flores, and M. B. Palma, "Redesign of the supply chain of a restaurant franchise in the food industry," *Global Journal of Business Research*, vol. 10, no. 2, pp. 103–111, 2016.
- [22] D. D. Zulfikar and D. Ernawati, "Pengukuran kinerja supply chain menggunakan metode green score DI PT," *XYZ. JUMINTEN*, vol. 1, no. 1, pp. 12–23, 2020.
- [23] G. Efreem, W. C. Mariam, A. Jemal, and H. Assefa, "Assessment of laboratory commodity supply chain system at public health

- facilities of Jimma zone and Jimma town administration, south west Ethiopia,” *International Journal of Research-GRANTHAALAYAH*, vol. 7, no. 10, pp. 471–490, 2019.
- [24] K. Anindita, I. G. A. A. Ambarawati, and R. K. Dewi, “Kinerja rantai pasok di pabrik gula madukismo dengan metode supply chain operation reference-analytical hierarchy process (SCOR-AHP),” *Agrisocionomics Jurnal Sosial Ekonomi Pertanian*, vol. 4, no. 1, pp. 125–134, 2020.
- [25] A. K. Joshi, I. A. Dandekar, and A. P. Shrotri, “Efforts taken by Indian administrative systems to manage the supply chain of essential commodities during the lockdown period of COVID-19-a case study,” *Journal of Information and Computational ence*, vol. 10, no. 4, pp. 804–811, 2020.
- [26] M. Shafiq, A. Akhtar, and A. H. Tahir, “A systematic review of efficiency and effectiveness in humanitarian organizations logistics and supply chain management (1995-2019),” in *Hamdard Islamicus: quarterly journal of the Hamdard National Foundation*, vol. 43no. 1, pp. 655–706, Pakistan, 2020.
- [27] T. A. Bui, H. T. T. Trinh, and B. T. Nguyen, “The potential for Tra Vinh province to become trade gateway of the Mekong Delta from logistics and supply chainmanagement perspectives,” *entific Journal of Tra Vinh University*, vol. 1, no. 4, pp. 23–40, 2020.
- [28] M. Ju and I. M. Kim, “ML performance analysis of the decode-and-forward protocol in cooperative diversity networks,” *IEEE Transactions on Wireless Communications*, vol. 8, no. 7, pp. 3855–3867, 2019.
- [29] L. Dwyer, T. Armenski, and L. K. Cvelbar, “Modified importance–performance analysis for evaluating tourism businesses strategies: comparison of Slovenia and Serbia,” *International Journal of Tourism Research*, vol. 18, no. 4, pp. 41–75, 2016.
- [30] W. Schulz, G. Diendorfer, and S. Pedebay, “The European lightning location system EUCLID – Part 1: performance analysis and validation,” *Natural Hazards and Earth System Sciences*, vol. 16, no. 2, pp. 595–605, 2016.
- [31] E. Frauman and S. Banks, “Gateway community resident perceptions of tourism development: incorporating importance–performance analysis into a limits of acceptable change framework,” *Tourism Management*, vol. 32, no. 1, pp. 128–140, 2017.
- [32] A. Aubry, V. Carotenuto, A. De Maio, A. Farina, and L. Pallotta, “Optimization theory-based radar waveform design for spectrally dense environments,” *IEEE Aerospace and Electronic Systems Magazine*, vol. 31, no. 12, pp. 14–25, 2017.
- [33] S. Balaji and C. V. Brown, “Lateral coordination mechanisms and the moderating role of arrangement characteristics in information systems development outsourcing,” *Information Systems Research*, vol. 25, no. 4, pp. 747–760, 2017.

Research Article

6G Green IoT Network: Joint Design of Intelligent Reflective Surface and Ambient Backscatter Communication

Qiang Liu ¹, Songlin Sun ¹, Heng Wang ², and Shaowei Zhang ²

¹School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China

²China Telecom Research Institute, Beijing, China

Correspondence should be addressed to Heng Wang; wangh26@chinatelecom.cn

Received 25 March 2021; Revised 24 May 2021; Accepted 4 June 2021; Published 22 June 2021

Academic Editor: Shengjie Xu

Copyright © 2021 Qiang 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.

Ambient backscatter communication (AmBC) is one of the candidate solutions for the 6G green internet of things (IoT) network. However, the uncontrollability of the radio frequency (RF) environment is one of the main obstacles hindering the popularization of AmBC. The intelligent reflective surface (IRS) can improve the radio frequency environment by adjusting the phase and amplitude of the incident signal, which provides the possibility for the widespread deployment of AmBC. Currently, there is no discussion about the joint optimization of AmBC and IRS. In this paper, we introduce a novel IRS and AmBC joint design method. The purpose of this method is to jointly design the beamforming vector, the IRS phase shift, and the reflection coefficient of AmBC to minimize the AP's transmit power while ensuring the quality of service of the AmBC system and the primary communication system. Due to the nonconvexity of the problem, the time complexity of solving the problem through exhaustive search will be very high. Therefore, we propose a joint design method based on an iterative beamforming vector, IRS phase shift, and reflection coefficient to minimize the AP's transmit power. This method can effectively reduce the transmission power of the access point (AP), and the simulation results prove the effectiveness of the method.

1. Introduction

With the rapid development of mobile communication technology, the internet of things (IoT) has been greatly developed and popularized. In particular, 5G/6G further promotes the application range of IoT, such as smart home, smart manufacturing, and smart cities [1, 2]. However, due to the increase in diversified requirements for the application scenarios of the IoT, diversified requirements are also put forward for the needs of the IoT devices. If the IoT device [3, 4] actively generates signals for wireless communication, it will consume a lot of energy, which will undoubtedly reduce the standby time of the device. Increasing the battery capacity will increase the standby time of the device, but this will undoubtedly increase the size and cost of the device. Especially for IoT devices such as wearable devices, they are very sensitive to device size and standby time. The battery capacity and size of IoT devices are the restrictive factors for their widespread popularity. Therefore, low-energy IoT

device transmission solutions are an important research direction to realize the potential of the IoT [3–7].

Radio frequency (RF) energy harvesting technology can obtain energy from external radio frequency sources and is one of the important research directions of low-energy consumption IoT device transmission solutions. RF energy harvesting technology has been widely used in low-power IoT devices. Wireless IoT devices can use RF energy harvesting technology to collect energy to maintain their normal operations. In this way, the wireless device can run for a long time without any manual intervention, thereby reducing the operation and maintenance costs of the device. Therefore, RF energy harvesting is particularly suitable for power-constrained wireless networks. There are three main types of RF energy harvesting schemes, including the synchronous wireless information and power transmission network (SWIPT), wireless power communication network (WPCN), and wireless power transmission (WPT) [8]. (1) The SWIPT scheme allows the transmitter to send information and

energy at the same time, and the user can choose to decode the information or collect energy. (2) The WPCN scheme allows user equipment to collect energy from RF energy signals and then actively send data. (3) The WPT scheme allows the power transmitter to transmit energy to the user equipment. Although these solutions have their application value in wireless networks, there are still some limitations. First, these solutions require a dedicated RF source to send RF energy or information to users. Secondly, active RF data transmission requires a complicated circuit design and consumes a lot of power.

As a green communication technology, ambient backscatter communication (AmBC) can effectively solve the above-mentioned limitations of traditional radio frequency energy harvesting technology [8, 9]. In the AmBC system, backscatter devices can communicate by using broadcast signals from RF sources such as cellular base stations, FM towers, and TV towers. In the AmBC system, the backscatter transmitter can modulate the data to the surrounding ambient signal and reflect it to the backscatter receiver. Therefore, AmBC does not need a dedicated frequency spectrum for data transmission. Therefore, AmBC has advantages that other communication methods do not have. First, AmBC does not require a dedicated spectrum for data transmission, which improves spectrum utilization. Secondly, since AmBC does not require a dedicated RF source, maintenance costs and deployment costs are reduced. These advantages can make AmBC widely used in many practical applications. AmBC has huge application potential in future low-energy scenarios, but it still faces many challenges. The quality of service (QoS) of AmBC is affected by factors such as the location of the RF, the type of RF, and the RF environment. Therefore, AmBC must be designed specifically for specific RF sources. In addition, to use ambient signals from licensed sources, the AmBC protocol must ensure that it does not interfere with the QoS of licensed users.

Intelligent reflective surfaces (IRSs) [10–12] can realize an intelligent and reconfigurable radio propagation environment for the B5G/6G wireless communication system [13–20]. The IRS is a plane containing a large number of low-cost passive reflective elements, each of which can independently change the phase and/or amplitude of the incident signal. The IRS can improve the required channel conditions, thereby achieving a substantial increase in wireless communication capacity and reliability. Intelligent reflective surfaces (IRSs) also have various practical advantages in implementation. First, compared with traditional active antenna arrays, IRS can only passively reflect impact signals without generating radio frequency resonance. Second, IRS does not have any noise amplification and self-interference. Third, due to the simple structure of the IRS, it can be easily deployed in any desired location. Finally, IRS has good compatibility and compatibility and can be integrated into existing communication systems.

There are many studies on AmBC or IRS [3–11], but there are no articles on the joint optimization design of AmBC and IRS. For example, [9] evaluated the performance of the environmental backscattering system but did not consider the role of IRS. Reference [10] used the IRS to enhance

the active communication system to achieve the goal of minimum transmission power. Reference [11] combined IRS beamforming and reflection design to enhance Bistatic Backscatter Networks. IRS is a means to optimize the performance of AmBC, so it is necessary to study the joint optimization design of IRS and AmBC. Therefore, in this article, we have conducted a joint optimization design for IRS and AmBC to ensure the quality of service of active communication and AmBC while minimizing the transmission power. The innovations of this paper are as follows:

- (1) We considered an IRS-assisted spectrum sharing system, where AmBC rides on the primary communication system. The receivers in the two systems are the same receiver and can demodulate the signals of the two systems. We call this receiver a cooperative receiver (CR). Specifically, after the CR demodulates the signal of the primary communication system, the signal of the AmBC is then demodulated based on the demodulated signal
- (2) Under the condition that both the main communication system and AmBC are constrained by the quality of service, we have studied the issue of the minimum transmit power of the access point (AP) based on IRS assistance. This problem is nonconvex, so convex optimization methods cannot be used directly to solve this problem. At the same time, to solve this problem through exhaustive search methods, the time complexity will be very high. Therefore, we propose an iterative optimization method to optimize the minimum transmit power of the AP. Through joint beamforming and IRS phase shift design, the proposed iterative optimization method can effectively reduce the minimum transmit power of the AP

The rest of this paper is organized as follows. Section 2 introduces system model and problem formulation. Section 3 presents the optimization algorithm based on an alternate iteration. Section 4 presents numerical results and Section 5 concludes the paper.

Notations: scalars are represented by italic letters, vectors are represented by bold lowercase letters, and matrices are represented by bold uppercase letters. $|x|$ represents the modulus of the complex number. $\|\mathbf{x}\|$ represents the Euclidean norm of the complex-valued vector \mathbf{x} . $\text{diag}(\mathbf{x})$ represents a diagonal matrix, and each diagonal item is a corresponding item in \mathbf{x} . $\text{tr}(\mathbf{X})$ represents the trace of the square matrix \mathbf{X} . $\mathbf{X} \succeq 0$ means \mathbf{X} is a positive semidefinite matrix.

2. System Model and Problem Formulation

2.1. System Setup. The intelligent reflective surface- (IRS-) enhanced spectrum sharing system includes a primary communication system and a secondary communication system, as shown in Figure 1. The primary communication system is a MISO downlink communication system, which consists of a receiver and an access point (AP) with M antennas. The secondary communication system is an AmBC system, which consists of a receiver and a backscatter device (BD).

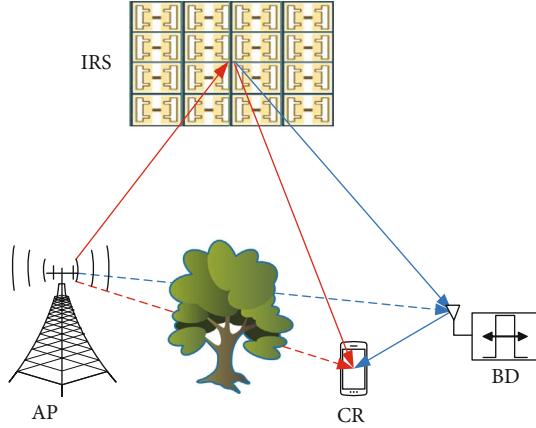


FIGURE 1: IRS-enhanced spectrum sharing system.

The backscatter device in the secondary system is a passive device, and its information transmission depends on the AP signal. We assume that the receiver in the primary communication system and the receiver in the secondary communication system are the same receiver. In other words, the receiver receives and demodulates the signal from the backscatter transmitter and the signal from the BS, simultaneously. For ease of expression, we denote the receiver as the cooperative receiver (CR). To improve QoS, an IRS with N passive reflective elements is used to help this spectrum sharing system communicate. The IRS equipped with an intelligent controller can be based on the signal propagation environment, and each reflective element can dynamically adjust the amplitude and phase shift of the incident signal.

2.2. IRS Model. IRS is a very promising green communication technology, which can reconfigure the wireless propagation environment through software. The IRS can modify the wireless channel between the transmitter and the receiver through a highly controllable reflection unit. This paved the way for the realization of a controllable wireless environment. Since IRS has no RF link, it has the advantages of low cost and low-energy consumption. Because the beam of the IRS is controllable, there is no need for complex interference management between IRSs. Assuming that the IRS is a frequency-selective surface, it allows certain RF signals to pass, absorb, or reflect certain signals. That is to say, IRS can reflect RF signals in a specific frequency band but cannot reflect RF signals in other frequency bands.

The IRS consists of N reflect elements, and each element $n \in \{1, 2, \dots, N\}$ can reflect the incident signal with a complex reflection coefficient. The complex reflection coefficient of the n th reflection element can be expressed as $\beta_n e^{j\theta_n}$, where $\beta_n \in [0, 1]$, $\forall n \in \{1, 2, \dots, N\}$ is the amplitude gain and $\theta_n \in [0, 2\pi)$, $\forall n \in \{1, 2, \dots, N\}$ is the phase shift. Although in theory, the amplitude gain can be adjusted within the interval $[0, 1]$. But adjusting amplitude gain and phase shift at the same time will greatly increase the complexity of the system. Therefore, without loss of generality, we take the upper bound of the interval $[0, 1]$ as the amplitude gain of all reflection elements, i.e., $\beta_n = 1$, $\forall n \in \{1, 2, \dots, N\}$. Then, the reflec-

tion coefficient matrix can be written as $\Theta = \text{diag}(e^{j\theta_1}, e^{j\theta_2}, \dots, e^{j\theta_N})$.

2.3. Backscatter Model. Since both IRS and BD can reflect signals, the signal will be reflected multiple times between IRS and BD, which greatly complicates the problem. We assume that the PT transmits a continuous wave signal with carrier frequency f_c and bandwidth B to communicate with the PR. In order to avoid the above-mentioned problems, BD adopts the following modulation method. First, BD uses a method similar to FSK modulation to shift the signal frequency f_c to frequency $f_c + \Delta f_c$ (only performs frequency shift; this process does not carry BD data) and then modulates the data that BD needs to send at frequency $f_c + \Delta f_c$, where Δf_c represents the frequency shift of the carrier frequency after BD modulation. Assume that the IRS can only reflect the signals with a specific frequency and bandwidth. Based on the difference in channel conditions, we assume that the IRS allows the reflection of the RF signal with carrier frequency f_c and bandwidth B but cannot reflect the RF signal with carrier frequency $f_c + \Delta f_c$ and bandwidth B . Therefore, to ensure that the signal sent by the PT and the signal reflected by the BD do not overlap in frequency, Δf needs to meet the constraint condition $\Delta f \geq B$. Although the above process occupies additional spectrum resources, it can effectively solve the problem of multiple reflections between IRS and BD.

In the process of ambient backscatter communication, we need to consider the power consumption constraints of the BD circuit; that is, the ambient signal energy received by the BD must meet the circuit power consumption constraints to activate the BD circuit for backscatter communication. Assuming that the minimum received signal power to maintain the normal operation of the BD circuit is P_{\min} . When the BD is semipassive, the BD needs other power sources to supply power. In this case, all received signals are used for reflection communication. When BD is passive, the energy of the input signal must be greater than P_{\min} . In this case, part of the received signal is used to power the BD circuit, and the other part is used for scatter communication. Since the passive BD has no power supply battery, its volume and cost have advantages compared with the semipassive BD. Therefore, in the following analysis, we mainly consider passive BD.

2.4. Transmission Model. We assume that the channel is flat fading and does not change during the coherence time. Then, we denote the channels of AP-IRS, BS-CR, AP-BD, AP-CR, IRS-CR, IRS-BD, and BD-CR as $\mathbf{H}_{ai} \in \mathbb{C}^{N \times M}$, $\mathbf{h}_{ac}^H \in \mathbb{C}^{1 \times M}$, $\mathbf{h}_{ab}^H \in \mathbb{C}^{1 \times M}$, $\mathbf{h}_{ci}^H \in \mathbb{C}^{1 \times N}$, $\mathbf{h}_{bi}^H \in \mathbb{C}^{1 \times N}$, and $h_{bc} \in \mathbb{C}^{1 \times 1}$. We assume that CR is assigned a linear beamforming vector, which can be denoted as $\mathbf{w} \in \mathbb{C}^{M \times 1}$. Then, the signal transmitted by AP is given as follows:

$$\mathbf{x}_{ap} = \mathbf{w}s, \quad (1)$$

where s is the signal that the main communication system needs to send and $E(|s|^2) = 1$. In this paper, we assume that IRS allows reflecting the RF signal with carrier frequency f_c

but cannot reflect the RF signal with carrier frequency $f_c + \Delta f_c$. Then, the signal received by the DB is mainly composed of two parts: one part is from AP and the other part is reflected by IRS. The signal received by BD can be expressed as

$$y_b = \left(\mathbf{h}_{bi}^H \Theta \mathbf{H}_{ai} + h_{ab}^H \right) w s. \quad (2)$$

Since no signal processing is performed in the BD, there is no noise term at (2), which is consistent with the backscatter literature. Since BD is a passive device, it needs to collect energy to power its circuit operation. Therefore, the signal received by the BD will be divided into two parts, which are used for circuit operation and signal reflection. Denote the reflection efficiency as α , then the α needs to satisfy the following constraint:

$$0 < \alpha \ll 1. \quad (3)$$

Let c denote the signal of BD, then the signal reflected by BD is given by

$$x_b = \sqrt{\alpha} y_b c e^{j2\pi \Delta f t}. \quad (4)$$

The remaining part is used to support the normal operation of the BD circuit. The power of the signal input for energy harvesting can be expressed as

$$P_b = (1 - \alpha) \eta \|y_b\|^2, \quad (5)$$

where η denotes the energy conversion efficiency of BD. Assume that the minimum power required to support the operation of the BD circuit is P_{\min} , then the following constraints should be satisfied:

$$(1 - \alpha) \eta \|y_b\|^2 \gg P_{\min}. \quad (6)$$

Denote the received signal of CR as $y_c(n)$, which is mainly composed of the signal from AP, IRS, and BD. Then, $y_c(n)$ is given by

$$y_c = \left(h_{ci}^H \Theta \mathbf{H}_{ai} + h_{ac}^H \right) w s(t) + h_{bc} x_b + n, \quad (7)$$

where $n \in CN(0, \sigma^2)$ denotes the Gaussian noise. Then, the received signal plus noise ratio (SNR) of demodulated $s(n)$ at the CR is given by

$$\gamma_s = \frac{\left| \left(\mathbf{h}_{ci}^H \Theta \mathbf{H}_{ai} + h_{ac}^H \right) w \right|^2}{\sigma^2}. \quad (8)$$

We assume that the primary communication system has a minimum SINR requirement and denote it as γ_{th}^p . Then, the QoS constraints of the primary communication system are given by

$$\gamma_s \gg \gamma_{th}^p. \quad (9)$$

After the CR successfully demodulates $s(n)$, the CR can decode the received signal $c(n)$ by performing successful interference cancellation (SIC). Then, the instantaneous received SNR of demodulated $c(n)$ at the CR is given by

$$\gamma_c = \frac{\alpha \left| \left(\mathbf{h}_{bi}^H \Theta \mathbf{H}_{ai} + h_{ab}^H \right) h_{bc} w \right|^2}{\sigma^2}. \quad (10)$$

We assume that AmBC has the minimum SNR requirement γ_{th}^a . To ensure the QoS of AmBC, the following conditions must be met:

$$\gamma_c \gg \gamma_{th}^a. \quad (11)$$

2.5. Problem Formulation. We study the issue of minimum transmit power under the condition that CR and BD meet their SNR requirements. Therefore, we need to jointly optimize the beamforming vector of AP, the phase shift of IRS, and the backscatter coefficient of BD to minimize transmit power of AP. Then, the corresponding optimization problem can be written as

$$(P1): \max_{\Theta, \alpha, w} \|w\|^2, \quad (12a)$$

$$\text{s.t. } (1 - \alpha) \eta \|y_b\|^2 \gg P_{\min}, \quad (12b)$$

$$\frac{\left| \left(h_{ci}^H \Theta \mathbf{H}_{ai} + h_{ac}^H \right) w \right|^2}{\sigma^2} \gg \gamma_{th}^p, \quad (12c)$$

$$\frac{\alpha \left| \left(\mathbf{h}_{bi}^H \Theta \mathbf{H}_{ai} + h_{ab}^H \right) h_{bc} w \right|^2}{\sigma^2} \gg \gamma_{th}^a, \quad (12d)$$

$$0 \ll \alpha \ll 1, \quad (12e)$$

$$0 \leq \theta_n \ll 2\pi, \quad \forall n = 1, 2, \dots, N. \quad (12f)$$

Obviously, (P1) is a nonconvex problem. There is no optimal solution to this problem. Next, we will analyze and simplify this problem so that it can be solved effectively.

3. Optimization Algorithm Based on Alternate Iteration

It can be seen that the problem (P1) is affected by multiple variables, which makes the problem difficult to solve. We use alternating optimization to solve problem (P1), which iteratively optimizes one variable while holding the others constant. In this section, we will introduce in detail how to solve problem (P1).

3.1. Transmit Beamforming Vector Optimization. When the phase of the IRS Θ and the reflection efficiency of the BD α are fixed, the vectors $\mathbf{h}_{ci}^H \Theta \mathbf{H}_{ai} + h_{ac}^H$, $\mathbf{h}_{bi}^H \Theta \mathbf{H}_{ai} + h_{ab}^H$, and $(\mathbf{h}_{bi}^H \Theta \mathbf{H}_{ai} + h_{ab}^H) h_{bc}$ are fixed. Define $g_{ac} = \mathbf{h}_{ci}^H \Theta \mathbf{H}_{ai} + h_{ac}^H$, $g_{ab} = \mathbf{h}_{bi}^H \Theta \mathbf{H}_{ai} + h_{ab}^H$, and $g_{bc} = (\mathbf{h}_{bi}^H \Theta \mathbf{H}_{ai} + h_{ab}^H) h_{bc}$, then problem (P1) can be expressed as

$$(P2): \min_w \|w\|^2, \quad (13a)$$

$$\text{s.t. } (1 - \alpha)\eta |g_{ab} w|^2 \gg P_{\min}, \quad (13b)$$

$$\frac{|g_{ac} w|^2}{\sigma^2} \gg \gamma_{th}^p, \quad (13c)$$

$$\frac{\alpha |g_{bc} w|^2}{\sigma^2} \gg \gamma_{th}^a. \quad (13d)$$

By observing (P2), we noticed that $\|w\|^2 = tr(w w^H)$, $|g_{ac} w|^2 = tr(w w^H g_{ac}^H g_{ac})$, and $|g_{ab} w|^2 = tr(w w^H g_{ab}^H g_{ab})$. Define $X = w w^H$, $G_{ac} = g_{ac}^H g_{ac}$, $G_{ab} = g_{ab}^H g_{ab}$, and $G_{bc} = g_{bc}^H g_{bc}$. Then, problem (P2) can be equivalently written as

$$(P3): \min_X tr(X), \quad (14a)$$

$$\text{s.t. } tr(X G_{ab}) (1 - \alpha)\eta \gg P_{\min}, \quad (14b)$$

$$tr(X G_{ac}) \gg \gamma_{th}^p \sigma^2, \quad (14c)$$

$$tr(X G_{bc}) \gg \gamma_{th}^a \sigma^2, \quad (14d)$$

$$X \succeq 0, \quad (14e)$$

$$\text{rank}(X) = 1, \quad (14f)$$

where problem (14a) is linear in X , then constraints (14b)–(14d) are linear inequalities in X . $X \succeq 0$ means that the matrix X is a symmetric positive semidefinite matrix, and the set of symmetric positive semidefinite matrices is convex. Note that the rank constraint in (14d) is the only nonconvex constraint. Therefore, we can use the SDR method to relax this constraint. Then, problem (P3) can be rewritten as

$$(P4): \min_X tr(X), \quad (15a)$$

$$\text{s.t. } tr(X G_{ab}) (1 - \alpha)\eta \gg P_{\min}, \quad (15b)$$

$$tr(X G_{ac}) \gg \gamma_{th}^p \sigma^2, \quad (15c)$$

$$tr(X G_{bc}) \gg \gamma_{th}^a \sigma^2, \quad (15d)$$

$$X \succeq 0. \quad (15e)$$

Obviously, problem (P4) is a standard convex semidefinite program (SDP), which can be optimized by a convex optimization solver such as CVX. Generally, the rank of the solution of problem (P4) is generally not equal to 1, which means that the optimal value of (P4) is the lower bound to satisfy (P3). Therefore, the solution of problem (P4) needs to be further processed to satisfy the constraint of problem (P3). First, we eigenvalue decomposition of X as $X = U \Sigma U^H$, where U is a unitary matrix and Σ is a diagonal matrix. Then, a suboptimal solution of problem (P3) can be expressed as $w = U \Sigma^{1/2} e$, where e is uniformly distributed on the unit sphere. w may not satisfy the constraints of (15b)–(15d). However, all constraints can be satisfied by simply scaling w to find a feasible weight vector.

3.2. IRS Phase Shift Optimization. Since the objective function (14a) in the problem (P1) depends only on w , the optimization of Θ can take the form of a feasibility problem. When w and α are given, problem (P1) can be expressed as

$$(P5): \text{find } \Theta, \quad (16a)$$

$$\text{s.t. } (1 - \alpha)\eta |g_{ab} w|^2 \gg P_{\min}, \quad (16b)$$

$$\frac{|g_{ac} w|^2}{\sigma^2} \gg \gamma_{th}^p, \quad (16c)$$

$$\frac{\alpha |g_{bc} w|^2}{\sigma^2} \gg \gamma_{th}^a, \quad (16d)$$

$$0 \leq \theta_n \ll 2\pi, \quad \forall n = 1, 2, \dots, N. \quad (16f)$$

Let $\mathbf{v} = [v_1, v_2, \dots, v_N]$, where $v_n = e^{j\theta_n}$, $\forall n = 1, 2, \dots, N$. Then, the constraints in (16f) are equivalent to $|v_n| = 1$, $\forall n = 1, 2, \dots, N$. In problem (P5), the variables related to Θ are g_{ab} , g_{ac} , and g_{bc} , so we need to change the forms of g_{ab} , g_{ac} , and g_{bc} to get the ideal expressions. By observing $g_{ab} = \mathbf{h}_{bi}^H \Theta \mathbf{H}_{ai} + h_{ab}^H$ and $g_{ac} = \mathbf{h}_{ci}^H \Theta \mathbf{H}_{ai} + h_{ac}^H$, g_{ab} and g_{ac} can be rewritten as $g_{ab} = \bar{\mathbf{h}}_{bi}^H \bar{\Theta} \bar{\mathbf{H}}_{ai}$ and $g_{ac} = \bar{\mathbf{h}}_{ci}^H \bar{\Theta} \bar{\mathbf{H}}_{ai}$, respectively, where $\bar{\mathbf{h}}_{bi}^H = [\mathbf{h}_{bi}^H \ 1]$, $\bar{\mathbf{h}}_{ci}^H = [\mathbf{h}_{ci}^H \ 1]$, $\bar{\Theta} = \text{diag}(e^{j\theta_1}, e^{j\theta_2}, \dots, e^{j\theta_N}, 1)$, and $\bar{\mathbf{H}}_{ai} = [\mathbf{H}_{ai} \ h_{ab}^H]$. Similarly, $g_{bc} = (\mathbf{h}_{bi}^H \Theta \mathbf{H}_{ai} + h_{ab}^H) h_{bc}$ can be rewritten as $g_{bc} = \bar{\mathbf{h}}_{ci}^H \bar{\Theta} \bar{\mathbf{H}}_{ai} h_{bc}$. Substituting $g_{ab} = \bar{\mathbf{h}}_{bi}^H \bar{\Theta} \bar{\mathbf{H}}_{ai}$ into (16b), we can get

$$(1 - \alpha)\eta \left| \bar{\mathbf{h}}_{bi}^H \bar{\Theta} \bar{\mathbf{H}}_{ai} w \right|^2 \gg P_{\min}. \quad (17)$$

Let $\Phi_{ab} = \text{diag}(\bar{\mathbf{h}}_{bi}^H \bar{\Theta} \bar{\mathbf{H}}_{ai}) w$ and $\bar{\mathbf{v}} = [\mathbf{v} \ 1]$, then $|\bar{\mathbf{h}}_{bi}^H \bar{\Theta} \bar{\mathbf{H}}_{ai} w|^2$ in (17) can be expressed as $|\bar{\mathbf{h}}_{bi}^H \bar{\Theta} \bar{\mathbf{H}}_{ai} w|^2 = |\bar{\mathbf{v}} \Phi_{ab}|^2$. Based on the following fact that $|\bar{\mathbf{v}} \Phi_{ab}|^2 = tr(\bar{\mathbf{v}} \Phi_{ab} \Phi_{ab}^H \bar{\mathbf{v}}^H) = tr(\bar{\mathbf{v}}^H \bar{\mathbf{v}} \Phi_{ab} \Phi_{ab}^H)$, we have $\mathbf{V} = \bar{\mathbf{v}}^H \bar{\mathbf{v}}$ and $\mathbf{R}_{ab} = \Phi_{ab} \Phi_{ab}^H$. Then, (17) can be rewritten as

$$tr(\mathbf{V} \mathbf{R}_{ab}) (1 - \alpha)\eta \gg P_{\min}. \quad (18)$$

Substituting $g_{ac} = \bar{\mathbf{h}}_{ci}^H \bar{\Theta} \bar{\mathbf{H}}_{ai}$ into (18), we can get

$$\left| \bar{\mathbf{h}}_{ci}^H \bar{\Theta} \bar{\mathbf{H}}_{ai} w \right|^2 \gg \gamma_{th}^p \sigma^2. \quad (19)$$

Let $\Phi_{ac} = \text{diag}(\bar{\mathbf{h}}_{ci}^H \bar{\Theta} \bar{\mathbf{H}}_{ai}) w$ and $\Phi_{bc} = \Phi_{ab} h_{bc}$, then $|\bar{\mathbf{h}}_{ci}^H \bar{\Theta} \bar{\mathbf{H}}_{ai} w|^2$ in (19) can be expressed as $|\bar{\mathbf{h}}_{ci}^H \bar{\Theta} \bar{\mathbf{H}}_{ai} w|^2 = |\bar{\mathbf{v}} \Phi_{ac}|^2$. Based on the following fact that $|\bar{\mathbf{v}} \Phi_{ac}|^2 = tr(\bar{\mathbf{v}}^H \bar{\mathbf{v}} \Phi_{ac} \Phi_{ac}^H)$, we can have $\mathbf{R}_{ac} = \Phi_{ac} \Phi_{ac}^H$. Then, (19) can be rewritten as

$$tr(\mathbf{V} \mathbf{R}_{ac}) \gg \gamma_{th}^p \sigma^2. \quad (20)$$

According to $|\bar{\mathbf{v}}\Phi_{bc}|^2 = \text{tr}(\bar{\mathbf{v}}^H\bar{\mathbf{v}}\Phi_{bc}\Phi_{bc}^H)$ and $\mathbf{V} = \bar{\mathbf{v}}^H\bar{\mathbf{v}}$, (16d) can be rewritten as

$$\text{tr}(\mathbf{V}\mathbf{R}_{ab})|h_{bc}|^2\alpha \gg \gamma_{th}^a\sigma^2. \quad (21)$$

Then, problem (P5) can be expressed as

$$(P6): \mathbf{find} \mathbf{V}, \quad (22a)$$

$$\mathbf{s.t.} \text{tr}(\mathbf{V}\mathbf{R}_{ab})(1-\alpha)\eta \gg P_{\min}, \quad (22b)$$

$$\text{tr}(\mathbf{V}\mathbf{R}_{ac}) \gg \gamma_{th}^p\sigma^2, \quad (22c)$$

$$\text{tr}(\mathbf{V}\mathbf{R}_{ab})|h_{bc}|^2\alpha \gg \gamma_{th}^a\sigma^2, \quad (22d)$$

$$\mathbf{V} \succeq \mathbf{0}, \quad (22e)$$

$$|v_n| = 1, \quad \forall n = 1, 2, \dots, N, \quad (22f)$$

$$\text{rank}(\mathbf{V}) = 1. \quad (22g)$$

To obtain an explicit solution of feasible \mathbf{V} , we can convert problem (P6) into an optimization problem based on constraints (22b)–(22f). Then, we introduce the slack variable μ to represent the difference between the achievable circuit constraint value and its requirement, then the optimization problem is as follows:

$$(P7): \mathbf{max}_{\mathbf{V}} \mu, \quad (23a)$$

$$\mathbf{s.t.} \text{tr}(\mathbf{V}\mathbf{R}_{ab})(1-\alpha)\eta \gg P_{\min} + \mu, \quad (23b)$$

$$\text{tr}(\mathbf{V}\mathbf{R}_{ac}) \gg \gamma_{th}^p\sigma^2, \quad (23c)$$

$$\text{tr}(\mathbf{V}\mathbf{R}_{ab})|h_{bc}|^2\alpha \gg \gamma_{th}^a\sigma^2, \quad (23d)$$

$$\mathbf{V} \succeq \mathbf{0}, \quad (23e)$$

$$|v_n| = 1, \quad \forall n = 1, 2, \dots, N, \quad (23f)$$

$$\text{rank}(\mathbf{V}) = 1. \quad (23g)$$

By relaxing the rank constraint, i.e., $\text{rank}(\mathbf{V}) = 1$, the problem can be transformed into a convex optimization problem, which can be easily solved using CVX. Then, we perform Gaussian randomization on the solution obtained by solving the CVX to obtain a rank-one solution. The process of Gaussian randomization is as follows. First, perform eigenvalue decomposition on \mathbf{V} to get $\mathbf{V} = \mathbf{U}\Sigma\mathbf{U}^H$, where $\mathbf{U} = [e_1, e_2, \dots, e_{N+1}]$ is a unitary matrix and $\Sigma = \text{diag}(\lambda_1, \lambda_1, \dots, \lambda_{N+1})$ is a diagonal matrix. Then, let $\bar{\mathbf{v}} = e^{j\arg(\mathbf{U}\Sigma^{1/2}r)}$, where $r \in \mathcal{CN}(0, 1)$. For the independently generated Gaussian random vector r , the target value $\bar{\mathbf{v}}$ is approximately the maximum one of problem (P7) among all r . Then, we can get the IRS phase shift $\mathbf{v} = e^{j[\bar{\mathbf{v}}\mathbf{v}[N+1]](1:N)}$, where $[\bar{\mathbf{v}}\mathbf{v}[N+1]](1:N)$ represents a vector containing the first N elements of $\bar{\mathbf{v}}\mathbf{v}[N+1]$.

3.3. Reflection Coefficient Optimization. Next, we will discuss the optimization method of reflection coefficient. The feasibility problem can be written as

$$(P8): \mathbf{find} \alpha, \quad (24a)$$

$$\mathbf{s.t.} (1-\alpha)\eta|g_{ab}w|^2 \gg P_{\min}, \quad (24b)$$

$$\frac{|g_{ac}w|^2}{\alpha|g_{bc}|^2 + \sigma^2} \gg \gamma_{th}^p, \quad (24c)$$

$$\frac{\alpha|g_{bc}w|^2}{\sigma^2} \gg \gamma_{th}^a, \quad (24d)$$

$$0 \ll \alpha \ll 1. \quad (24e)$$

Based on (24b)–(24e), we can get the value range of α . The value range of α can be expressed as follows:

$$\frac{\sigma^2\gamma_{th}^a}{|g_{bc}w|^2} \ll \alpha \ll \min\left(1 - \frac{P_{\min}}{\eta|g_{ab}w|^2}, \frac{|g_{ac}w|^2}{\gamma_{th}^p|g_{bc}|^2} - \frac{\sigma^2}{|g_{bc}|^2}\right). \quad (25)$$

Obviously, the value of α is determined by w and Θ . Therefore, if the optimal values of w and Θ cannot be determined, we cannot obtain the optimal value of α in a limited time.

However, by observing Algorithms 1 and 2, we can draw the following conclusions: for given α and Θ , we can find the optimal w of problem (P2) according to Algorithm 1; for given α and w , we can find the optimal Θ of problem (P5) according to Algorithm 2. Therefore, for a given α , we can alternately use Algorithms 1 and 2 to solve the local optimal solution of problem (P1). To facilitate practical implementation, we consider that the backscattering coefficient can only adopt a limited number of discrete values. Let L indicate the number of backscattering coefficient levels. For simplicity, we assume that such discrete backscattering coefficients are obtained by uniformly quantizing the interval $(0, 1]$. Thus, the set of discrete backscattering coefficients is given by

$$\mathcal{F} = \{\Delta\alpha, 2\Delta\alpha, \dots, 1\}, \quad (26)$$

where $\Delta\alpha = 1/L$. Let $\alpha_l = l\Delta\alpha, l = 1, 2, 3, \dots, L$. Then, we can solve the local optimal values Θ_l and w_l for each α_l of problem (P1). Let $W = [\|w_1\|^2, \|w_2\|^2, \dots, \|w_L\|^2]$. Then, the optimal solution of problem (P1) can be given by

$$(w^*, \Theta^*, \alpha^*) = \arg \min (W). \quad (27)$$

Then, take the Θ_l, w_l , and α_l that minimize $\|w_l\|^2$ as the solution to problem (P1). Algorithm 3 gives a detailed description of the alternate optimization algorithm, where ε is a threshold for the increment of the objective value until convergence.

4. Simulation Results

In this section, we will evaluate the performance of the algorithm. To effectively evaluate the proposed algorithm, we consider the settings shown in Figure 2. The default locations

1:**Initialize:** random IRS phase shifts Θ ; random backscatter coefficients α ;
 2:Optimize problem (P4) by CVX and get X .
 3:Get U and Σ , where $X = U\Sigma U^H$
 4:Get $w = U\Sigma^{1/2}e$, where e is uniformly distributed on the unit sphere.
 5:Scaling w so as to satisfy constraints (14b)–(14d).

ALGORITHM 1: Transmit beamforming vector optimization.

1:**Initialize:** backscatter coefficients α generated in Algorithm 1, transmit beamforming vector w generated in Algorithm 1; number of Gaussian randomization G
 2:Relax the constraint (23d), then optimize problem (P7) by CVX and get \mathbf{V} .
 3:Perform eigenvalue decomposition $\mathbf{V} = U\Sigma U^H$
 4:**for** $i = 1$ to G **do**
 5:Get a rank-one solution of (P7): $\bar{\mathbf{v}}_i = e^{j\arg(U\Sigma^{1/2}r)}$ **end for**
 6:Obtain the target value of problem (P7), where $\bar{\mathbf{v}} = \max \mu(\bar{\mathbf{v}}_i)$
 7:Obtain the target value of problem (P5), where $\mathbf{v} = e^{j[\bar{\mathbf{v}}/v_1[N+1]](1:N)}$

ALGORITHM 2: IRS phase shift optimization.

1:**Initialize:** number of backscattering coefficient levels L , number of Gaussian randomization G , and threshold ϵ .
 2:**for** $l = 1$ to L **do**
 3:random IRS phase shifts Θ_l
 4:**While** the change of the objective function (12a) is higher than the threshold ϵ **do**
 5:Optimize problem (P4) by CVX and get X .
 6:Get U and Σ , where $X = U\Sigma U^H$
 7:Get $w_l = U\Sigma^{1/2}e$, where e is uniformly distributed. On the unit sphere.
 8:Scaling w_l so as to satisfy constraints (14b)–(14d).
 9:Relax the constraint (23d), then optimize problem (P7) by CVX and get \mathbf{V} .
 10:Perform eigenvalue decomposition $\mathbf{V} = U\Sigma U^H$
 11:**for** $i = 1$ to G **do**
 12:Get a rank-one solution of (P7), $\bar{\mathbf{v}}_i = e^{j\arg(U\Sigma^{1/2}r)}$.
 13:Obtain the target value of problem (P7), where $\bar{\mathbf{v}}_l = \arg \max \mu(\bar{\mathbf{v}}_i)$.
 14:Obtain the target value of problem (P5), where $\mathbf{v}_l = e^{j[\bar{\mathbf{v}}_l/v_1[N+1]](1:N)}$.
 15:Set $\Theta_l = \text{diag}(\mathbf{v}_l)$ for the next iteration.
 16: **end while**
 17: $(w^*, \Theta^*, \alpha^*) = \arg \min (\|w_l\|^2)$.
 18:**end for**
 19:**Return:** optimized beamforming vector w^* , optimized phase shift vector Θ^* , optimized reflection coefficient α^* .

ALGORITHM 3: The alternate optimization algorithm.

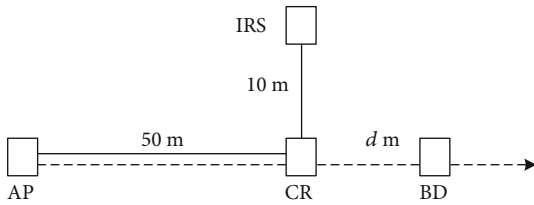


FIGURE 2: Simulation setup of the IRS-aided IRS spectrum sharing system.

of the AP, IRS, CR, and BD are $(0, 0)$, $(50, 10)$, $(50, 0)$, and $(d, 0)$, with all coordinates in meters hereafter.

The default number of IRS elements is $N = 100$, while the AP has 8 antennas. We assume that all channels are independent Rayleigh fading, and the path loss index is set to 2.2 and the reference distance is 1 m. For all channels, the path loss at 1 meter (m) is set to 30 dB. For ease of analysis, we assume that each channel coefficient is uniformly randomly generated from $[0, 2\pi)$. Since there is occlusion between AP and CR (BD), we assume that the penetration loss is 10 dB. At the same time, we assume that the antenna gain of AP, BD, and CR are all 0 dBi and the antenna gain of each reflective

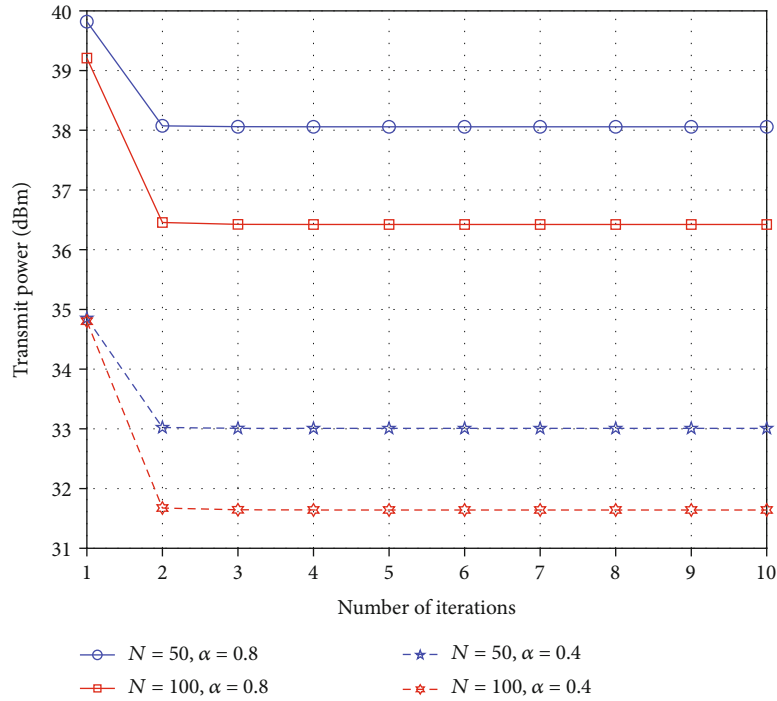


FIGURE 3: Convergence of the proposed distributed algorithm.

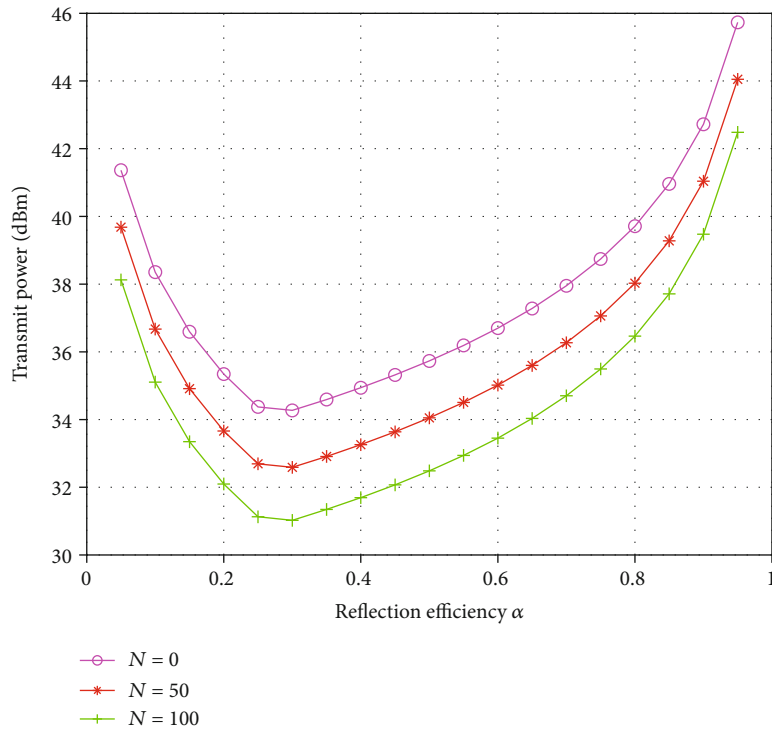


FIGURE 4: The influence of the number of reflection units.

element of the IRS is 5 dBi. The noise power is set to -90 dBm. The threshold ε is set to 0.01. We set the minimum SNR required to demodulate the primary signal and the BD signal to 20 dB and 13 dB, respectively.

First, we verified the convergence of the algorithm. When verifying the convergence of the algorithm, we do not con-

sider the impact of the threshold on the algorithm but only consider the impact of the number of iterations on the algorithm. To qualitatively analyze the convergence of the algorithm, we locate BD at (52, 0). At the same time, in order to illustrate the influence of the number of IRS units on the convergence of the algorithm, we considered two cases where

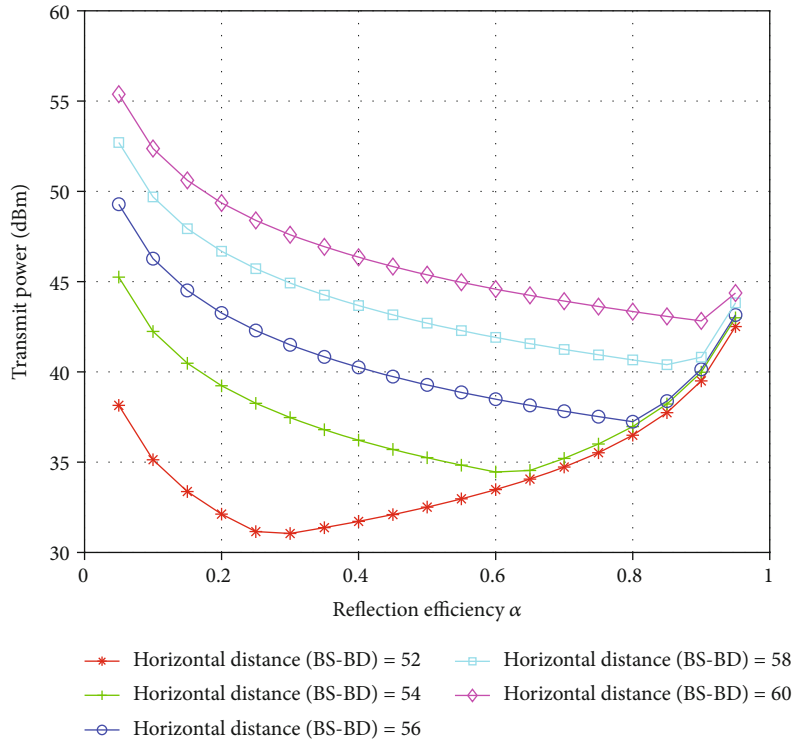


FIGURE 5: The influence of horizontal distance (BS-BD).

the number of IRS elements is $N = 50$ or $N = 100$. As shown in Figure 3, as the number of iterations increases, the AP's transmit power gradually decreases and tends to stabilize. This shows that the proposed algorithm has good convergence. As shown in Figure 3, when considering the impact of the threshold on the proposed algorithm, the transmit power can converge at most in three iterations. That is to say, when the maximum number of iterations is set to 3, the proposed algorithm can obtain satisfactory transmit power. Since the complexity of the algorithm is proportional to the number of iterations, the more iterations, the lower the complexity. Therefore, Figure 3 also proves that the complexity of the proposed algorithm is low. At the same time, we can see from Figure 3 that the greater the number of IRS units, the smaller the transmit power required by the AP. And the performance of $\alpha = 0.4$ is better than the performance of $\alpha = 0.8$.

Although Figure 3 shows the effect of different reflection efficiencies on system performance, it does not reflect the optimal reflection efficiency required by the proposed algorithm. In Figure 4, we show the effect of different reflection efficiencies on AP transmit power. We can see that although increasing the number of IRS units can reduce the transmission power, the optimal reflection coefficient setting has nothing to do with the number of reflection units. At the same time, we can see that the performance of the proposed algorithm is better than that of the system without IRS assistance.

However, it can be seen from Figure 5 that the choice of the best reflection coefficient is related to horizontal distance between AP and BD. We can see from Figure 5 that as the

horizontal distance between AP and BD increases, the optimal emission coefficient increases. We can also see from Figure 5 that as the horizontal distance between AP and BD increases, the transmission power required to ensure the QoS also increases.

5. Conclusions

Intelligent reflector surfaces (IRSs) can improve the radio frequency environment by adjusting the phase and amplitude of the incident signal, which provides the possibility for the widespread deployment of AmBC. In this paper, we introduce a novel IRS and AmBC joint design method. This method is based on the joint design of an iterative beamforming vector, IRS phase shift, and reflection coefficient to minimize the AP's transmit power. This method can effectively reduce the transmission power of the access point, and the simulation results prove the effectiveness of this method.

Data Availability

The data in this paper is based on MATLAB simulation. According to the method described in this paper, all the data can be obtained through MATLAB.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Key R&D Program of China (No. 2020YFB1806700).

References

- [1] Z. Zhang, Y. Xiao, Z. Ma et al., "6G wireless networks: vision, requirements, architecture, and key technologies," *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 28–41, 2019.
- [2] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: applications, trends, technologies, and open research problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, 2020.
- [3] A. N. Parks, A. Liu, S. Gollakota, and J. R. Smith, "Turbocharging ambient backscatter communication," in *Proceedings of the 2014 ACM Conference on SIGCOMM*, pp. 619–630, Chicago, IL, USA, August 2014.
- [4] K. Han and K. Huang, "Wirelessly powered backscatter communication networks: modeling, coverage, and capacity," *IEEE Transactions on Wireless Communications*, vol. 16, no. 4, pp. 2548–2561, 2017.
- [5] W. Liu, K. Huang, X. Zhou, and S. Durrani, "Full-duplex backscatter interference networks based on time-hopping spread spectrum," *IEEE Transactions on Wireless Communications*, vol. 16, no. 7, pp. 4361–4377, 2017.
- [6] D. T. Hoang, D. Niyato, P. Wang, D. I. Kim, and Z. Han, "Ambient backscatter: a new approach to improve network performance for RF-powered cognitive radio networks," *IEEE Transactions on Communications*, vol. 65, no. 9, pp. 3659–3674, 2017.
- [7] X. Lu, D. Niyato, H. Jiang, D. I. Kim, Y. Xiao, and Z. Han, "Ambient backscatter assisted wireless powered communications," *IEEE Wireless Communications*, vol. 25, no. 2, pp. 170–177, 2018.
- [8] N. van Huynh, D. T. Hoang, X. Lu, D. Niyato, P. Wang, and D. I. Kim, "Ambient backscatter communications: a contemporary survey," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 2889–2922, 2018.
- [9] X. Kang, Y.-C. Liang, and J. Yang, "Riding on the primary: a new spectrum sharing paradigm for wireless-powered IoT devices," *IEEE Transactions on Wireless Communications*, vol. 17, no. 9, pp. 6335–6347, 2018.
- [10] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5394–5409, 2019.
- [11] X. Jia, J. Zhao, X. Zhou, and D. Niyato, "Intelligent reflecting surface-aided backscatter communications," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, Taipei, Taiwan, December 2020.
- [12] X. Yu, D. Xu, and R. Schober, "MISO wireless communication systems via intelligent reflecting surfaces," in *2019 IEEE/CIC International Conference on Communications in China (ICCC)*, pp. 735–740, Changchun, China, August 2019.
- [13] M. M. Rana and M. M. R. Shuvo, "Localization of sensor nodes using weighted least squares algorithm: comprehensive literature review and future research directions," in *2018 International Conference on Innovation in Engineering and Technology (ICIET)*, pp. 1–6, Dhaka, Bangladesh, December 2018.
- [14] M. M. Rana, J. Kim, and W.-K. Cho, "Performance analysis of sub-carrier mapping in LTE uplink systems," in *Digest of the 9th International Conference on Optical Internet (COIN 2010)*, pp. 1–3, Jeju, Korea (South), July 2010.
- [15] S. Sun, M. Kadoch, L. Gong, and B. Rong, "Integrating network function virtualization with SDR and SDN for 4G/5G networks," *IEEE Network*, vol. 29, no. 3, pp. 54–59, 2015.
- [16] Y. Wu, B. Rong, K. Salehian, and G. Gagnon, "Cloud transmission: a new spectrum-reuse friendly digital terrestrial broadcasting transmission system," *IEEE Transactions on Broadcasting*, vol. 58, no. 3, pp. 329–337, 2012.
- [17] B. Rong, Y. Qian, K. Lu, H.-H. Chen, and M. Guizani, "Call admission control optimization in WiMAX networks," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 4, pp. 2509–2522, 2008.
- [18] N. Chen, B. Rong, X. Zhang, and M. Kadoch, "Scalable and flexible massive MIMO precoding for 5G H-CRAN," *IEEE Wireless Communications*, vol. 24, no. 1, pp. 46–52, 2017.
- [19] B. Rong, Y. Qian, and K. Lu, "Integrated downlink resource management for multiservice WiMAX networks," *IEEE Transactions on Mobile Computing*, vol. 6, no. 6, pp. 621–632, 2007.
- [20] S. Sun, L. Gong, B. Rong, and K. Lu, "An intelligent SDN framework for 5G heterogeneous networks," *IEEE Communications Magazine*, vol. 53, no. 11, pp. 142–147, 2015.